

Open Research Online

The Open University's repository of research publications and other research outputs

Characterization of CONSTANS, an Arabidopsis gene that promotes flowering.

Thesis

How to cite:

Robson, Frances Clare (1999). Characterization of CONSTANS, an Arabidopsis gene that promotes flowering. PhD thesis The Open University.

For guidance on citations see [FAQs](#).

© 1998 Frances Clare Robson



<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Version: Version of Record

Link(s) to article on publisher's website:

<http://dx.doi.org/doi:10.21954/ou.ro.00010217>

Copyright and Moral Rights for the articles on this site are retained by the individual authors and/or other copyright owners. For more information on Open Research Online's data [policy](#) on reuse of materials please consult the policies page.

oro.open.ac.uk

UNRESTRICTED

**Characterization of *CONSTANS*, an
Arabidopsis gene that promotes
flowering**

FRANCES CLARE ROBSON BSc

**A thesis submitted in partial fulfilment of the requirements
of the
Open University for the degree of Doctor of Philosophy**

June 1998

John Innes Centre

AUTHOR'S NUMBER: P9274776 Norwich

DATE OF AWARD: 19 FEBRUARY 1999 UK

ProQuest Number: 27696791

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 27696791

Published by ProQuest LLC (2019). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Abstract

Under long day conditions *Arabidopsis thaliana* plants grow vegetatively for approximately two weeks before they initiate flowering while under short day conditions the vegetative period lasts around six weeks. The *constans* (*co*) mutant flowers later than wild type under long days but at approximately the same time as wild type under short days. This suggests that the *CO* product promotes flowering in response to long days. The *CO* gene was cloned previously by chromosome walking and encodes a 373 amino acid protein containing two zinc fingers suggesting that it binds DNA and is involved in regulation of gene expression.

A series of mutant alleles of the *CO* gene was characterized and the effect of each mutation on flowering time examined. All seven *co* mutants are semi-dominant, with the heterozygotes having flowering times and leaf numbers intermediate between wild type and the homozygous *co* parent, and all differ quantitatively with respect to flowering time and leaf number.

The *co* alleles were characterized at the molecular level and the mutations were clustered into two regions of the gene that define important domains of the predicted *CO* protein. Five of the mutations affect the zinc finger region near the N-terminus of the protein and the remaining two mutations affect adjacent amino acids in a basic domain near the C-terminus of the protein. Both of these domains and the individual amino acids affected by the mutations are highly conserved. Homology searches revealed a family of at least 14 closely related genes in *Arabidopsis* and homologues in a number of other plant species.

The seven *co* mutations all cause in-frame changes and none were unambiguous null alleles. To investigate the effect of loss of *CO* function a putative null allele, *co-8*, was isolated after gamma-irradiation. This was characterized at the molecular level and found to contain a 1.3 kb deletion which removes approximately 1 kb of the promoter and 0.3 kb of the coding sequence at the 5' end of the *CO* ORF that includes the zinc finger domain. The new *co-8* mutant has an intermediate flowering time with respect to the seven previously isolated *co* mutants, and is also semi-dominant.

A number of constructs designed to express derivatives of the *CO* protein were introduced into *Arabidopsis* to investigate whether it was possible to repress flowering by the creation of dominant negative mutations. However, with the exception of one or two small effects these did not greatly affect flowering time.

A *Brassica napus* homologue of the *CO* gene was isolated. The predicted protein is 74% identical to the *Arabidopsis* *CO* protein, with the greatest conservation over the zinc fingers (91% identical) and C-terminal basic domain (94% identical). The homologue was able to complement the *Arabidopsis* *co-2* mutation and restore wild type flowering time.

Acknowledgements

I would like to thank the following people for their help and support during my PhD:-

My supervisor George Coupland for his help and advice during my time in his group, for accepting and encouraging my progression from assistant to PhD student to post-doc, and in particular for taking the time to thoroughly (and constructively) criticize my thesis as it took shape.

Jo Putterill for her guidance and supervision during the early years and for her support since the beginning of my PhD.

All members of the Coupland group, past and present, but in particular Karen Lee, Jon Clarke, Dave Stevenson, Paul Reeves, Hitoshi Onouchi, Kay Wheatley, Kathryn Graves, Paula Suárez-López, Giovanni Murtas, Alon Samach, Manolo Pineiro, Manuela Costa and Shelley Hepworth who have all been around during my writing up phase, some for a long time before and some newly arrived. Thank you for your friendship, for always being ready for a coffee, and for putting up with my mood swings. I really appreciate it. Past members of the Coupland group; Robert Schaffer, Maribel Igeño, Kate Wilson, Kate Dixon and Laurian Robert, for e-mailing at regular intervals. My other friends in Molecular Genetics, in particular Clare Lister, Yaron Levy, Tony Gendall, Ruth Meissner, Fred Rook and Claire Hutchison for friendship and company during lunch and Claus Schwechheimer and Kati King for explaining the mysterious workings of computers to me. Cindy Cooper and Jill Murton for making us laugh and for keeping the media kitchen in Molecular Genetics running smoothly. Xiaohu Hu for her technical assistance over the past year.

Phil Smith for his help with the statistical analysis, and for his companionship while we wrote up together in the library. My ex-flatmate Helma Tatge and my current flatmate Vibeke Alm, for their friendship, for feeding the cats and for not minding me moaning to them about writing up.

Matt Hills for his support and encouragement, for cooking endless meals, providing unlimited supplies of alcohol and chocolate and for being so patient and understanding.

The library staff for allowing me to live on the premises for so long and in particular Ingrid Walton for the translation of the words of Johannes Thal and Rachel Lewis for always being cheerful and friendly and for photocopying references for me when time was short. The greenhouse staff, in particular Mervyn who has worked miracles in S67 since he took over the care of our plants. The photography department, in particular Andrew Davis, who took most of the plant photographs in this thesis. The computing department, in particular Bob Findlay for patiently helping me with the thesis template program and Doug Atkinson for helping to find and set up my computer. The security guards for making me cups of coffee.

My parents for their wholehearted and much appreciated support and encouragement. My sister Charlotte and my brother-in-law, Michael, for cheerfully providing a holiday retreat on so many occasions over the past few years.

And Beryl and Soot, for being there.

Abstract

Acknowledgements

Abbreviations 1

1	Introduction	3
1.1	Flowering	3
1.1.1	Control of flowering by photoperiod	4
1.2	<i>Arabidopsis thaliana</i>	6
1.2.1	Flowering in <i>Arabidopsis thaliana</i>	10
1.2.1.1	Environmental control of flowering in <i>Arabidopsis</i>	13
1.2.1.1.1	Photoperiod and light quality	13
1.2.1.1.2	Vernalization	17
1.2.1.2	Genetic analysis of the control of flowering time	19
1.2.1.2.1	Natural variation in flowering time	19
1.2.1.2.2	Mutations affecting flowering time	21
1.2.1.2.2.1	Late-flowering mutants	22
1.2.1.2.2.2	Early-flowering mutants	24
1.2.1.2.3	Models for the genetic control of flowering time	25
1.2.1.2.3.1	The environmental response pathway	27
1.2.1.2.3.2	The autonomous pathway	29
1.2.1.2.3.3	The gibberellic acid promotion pathway	30

1.2.1.3	The role of <i>CO</i> in the promotion of flowering	31
1.3	Isolation of the <i>CONSTANS</i> gene	34
1.4	The <i>CONSTANS</i> protein has a novel cysteine-rich domain which is predicted to be able to form two zinc fingers	35
1.4.1	Zinc finger proteins	36
1.4.1.1	Comparison of the <i>CO</i> protein with other zinc finger proteins	41
1.5	Background and summary	42
2	Characterization of the phenotypes caused by seven <i>constans</i> mutant alleles	43
2.1	Introduction	43
2.1.1	Backcrossing <i>co-1</i> into Landsberg <i>erecta</i>	43
2.2	The phenotypes caused by seven <i>co</i> mutant alleles	45
2.2.1	Flowering time in long days	45
2.2.2	Flowering time in short days	46
2.2.3	Other phenotypes	47
2.2.3.1	Chimeric flowering shoots	47
2.2.3.2	Altered number of floral organs	48
2.3	The semi-dominance of the <i>constans</i> mutations	49
2.4	Discussion	50
3	Molecular characterization of the <i>constans</i> alleles	55
3.1	Introduction	55
3.2	Identifying the mutations in the seven <i>constans</i> alleles	55
3.2.1	Southern analysis	55
3.2.2	PCR amplification and cloning of the <i>constans</i> alleles	56
3.2.3	Sequencing the <i>constans</i> alleles	57
3.2.4	The mutations in the seven <i>constans</i> alleles define domains of the protein that are important for <i>CO</i> function	58
3.2.5	Analysis of expression of <i>co</i> mutant alleles	61

3.3	Homology to other genes	62
3.4	Discussion	67
4	Isolation and characterization of <i>co-8</i>	80
4.1	Introduction	80
4.2	Screening for new <i>co</i> alleles in an M ₂ population made from gamma-irradiated pollen	82
4.2.1	Identification of a new allele, <i>co-8</i>	83
4.3	Characterization of <i>co-8</i>	84
4.3.1	Molecular characterization of <i>co-8</i>	84
4.3.1.1	Analysis of <i>co-8</i> by Southern hybridization	84
4.3.1.2	Construction of a <i>co-8</i> cosmid library ...	86
4.3.1.2.1	Screening the library	86
4.3.1.2.2	Characterization of <i>co-8</i> cosmid clones by restriction mapping, Southern hybridization and subcloning	86
4.3.1.3	Isolation of the DNA flanking the <i>CO</i> ORF in <i>co-8</i> by IPCR	87
4.3.1.4	The deletion at the <i>CO</i> locus in <i>co-8</i> ...	88
4.3.1.5	Confirmation of the rearrangement in <i>co-8</i> by PCR	88
4.3.1.6	Determining the chromosomal location of the second breakpoint in <i>co-8</i>	90
4.3.1.7	Analysis of RNA produced by the <i>co-8</i> allele	91
4.3.1.8	Sequence analysis of the <i>co-8</i> rearrangement	91
4.3.1.9	A summary of the rearrangement in the <i>co-8</i> allele	92
4.3.2	Phenotypic characterization of <i>co-8</i> mutants	93
4.3.2.1	Flowering time	93
4.3.2.2	Other phenotypes	94

4.3.2.3	<i>co-8</i> is semidominant	95
4.4	Discussion	95
5	Overexpression of modified CONSTANS proteins	102
5.1	Introduction	102
5.2	Description of the modified CONSTANS proteins	103
5.3	Plasmid construction	104
5.3.1	<u>35S:<i>co-7</i></u>	104
5.3.2	<u>35S:<i>co-3</i></u>	105
5.3.3	<u>35S:ZF (Zinc Fingers)</u>	105
5.4	Introduction of the constructs into <i>Arabidopsis thaliana</i>	107
5.5	Identification of homozygous families carrying the T-DNA at a single locus	108
5.6	Phenotypic analysis of the transformants under long and short days	108
5.6.1	<u>35S:<i>co-7</i></u>	109
5.6.2	<u>35S:<i>co-3</i></u>	111
5.6.3	<u>35S:ZF</u>	111
5.7	Analysis of the expression of the transgenes by Northern hybridization	112
5.8	Discussion	112
6	Investigation of the conserved function of a <i>Brassica napus</i> CO homologue	120
6.1	Introduction	120
6.2	Sequence comparison of the <i>Brassica napus</i> CO homologue <i>Bn1CON10</i> , and <i>Arabidopsis thaliana</i> CO	123
6.3	Plasmid construction	123
6.3.1	The <i>Brassica napus</i> CO homologue expressed from its native promoter	124
6.3.2	The <i>Arabidopsis</i> CO promoter with the <i>Brassica napus</i> CO coding sequence	124
6.4	Introduction of the constructs into the <i>co-2</i> mutant	125

6.5	Identification of transgenic plants homozygous for the T-DNAs	126
6.6	Phenotypic analysis of the transformants under long and short days	126
6.6.1	Flowering time of transformants carrying BnCO	126
6.6.1.1	Phenotypes other than early flowering conferred by pBnCO	127
6.6.2	Flowering time of transformants carrying ApBnCO	128
6.7	Analysis of the expression of the transgenes by Northern hybridization and RT-PCR	128
6.8	Discussion	130
7	Materials and Methods	137
7.1	Materials	137
7.1.1	Plant Materials	137
7.1.2	Bacterial and Yeast Strains	137
7.1.3	Plasmids	138
7.1.4	General Media and Solutions	139
7.1.5	Plant Media	140
7.2	Methods	142
7.2.1	Bacterial and General DNA Techniques	142
7.2.1.1	Methods of DNA Extraction	142
7.2.1.2	Glycerol Stocks	143
7.2.1.3	Preparation of Competent <i>E. coli</i> Cells ..	144
7.2.1.4	Transformation of <i>E. coli</i>	144
7.2.1.5	Transfer of Plasmids from <i>E. coli</i> to <i>Agrobacterium tumefaciens</i>	145
7.2.1.6	Preparation of Filters for Colony Hybridization	146
7.2.1.7	Cleavage of DNA with Restriction Enzymes	147
7.2.1.8	Agarose Gel Electrophoresis of DNA ...	148
7.2.1.9	Isolation and Recovery of DNA Fragments from Agarose Gels	149

7.2.1.10	DNA Ligations	149
7.2.1.11	Southern Blotting	150
7.2.1.12	Hybridization of Radio-labelled DNA probes to DNA Immobilized on Nylon Membranes	151
7.2.1.13	Autoradiography	152
7.2.1.14	Radioactive Labelling of DNA Fragments	152
7.2.1.15	Quantification of Nucleic Acid Concentration	153
7.2.1.16	Ethanol Precipitation of DNA	153
7.2.1.17	Isopropanol Precipitation of DNA	154
7.2.1.18	Phenol/Chloroform Extraction of DNA . .	154
7.2.2	Plant DNA Extraction Methods	154
7.2.2.1	CTAB (Cetyl trimethyl ammonium bromide) Plant DNA Extraction	154
7.2.2.2	Dellaporta DNA Microprep	155
7.2.3	Polymerase Chain Reaction Techniques	156
7.2.3.1	PCR Amplification of Genomic and Plasmid DNA	156
7.2.3.2	Rapid Preparation of Plant Material for Diagnostic Screening by PCR	157
7.2.3.3	Inverse Polymerase Chain Reaction (IPCR)	157
7.2.3.4	Cloning PCR Products	158
7.2.3.5	Colony PCR	158
7.2.3.6	Oligonucleotide Preparation	159
7.2.4	RNA Extraction and Manipulation	161
7.2.4.1	Small Scale Total RNA Extraction from Plant Tissue	161
7.2.4.2	Northern Analysis	162
7.2.4.3	RT-PCR (Reverse transcription - Polymerase Chain Reaction)	163
7.2.5	DNA Sequencing	164

7.2.5.1	Sequencing Reactions	164
7.2.5.2	Polyacrylamide Gel Electrophoresis	165
7.2.5.3	Analysis of DNA Sequences	165
7.2.6	The Preparation and Screening of a Cosmid Library	166
7.2.7	Yeast Techniques	169
7.2.7.1	Screening the CIC YAC (Yeast Artificial Chromosome) Library	169
7.2.7.2	Yeast Genomic DNA Extraction	170
7.2.8	<i>Arabidopsis</i> Transformation and Tissue Culture Techniques	171
7.2.8.1	Transformation of <i>Arabidopsis thaliana</i> Root Explants	171
7.2.8.2	Transformation of <i>Arabidopsis thaliana</i> by Vacuum Infiltration	172
7.2.8.3	Seed Sterilization	174
7.2.8.4	Selective Assay on Kanamycin	174
7.2.9	Plant Growth Conditions	175
Appendix		177
References		191

Figures

- Figure 1.1** A diagram to illustrate possible interactions of the pathways that control the transition to flowering in *Arabidopsis*.
- Figure 1.2** Comparison of the amino acid sequences of the two predicted zinc fingers within the CO sequence.
- Figure 2.1** Wild type Landsberg *erecta* and the *co-2* mutant grown in long day (LD) and short day (SD) conditions.
- Figure 2.2** The *co-1* mutant in Landsberg with the wild type *ERECTA* gene compared to *co-1* in Landsberg *erecta*.
- Figure 2.3** The flowering times, represented as the mean rosette and cauline leaf numbers of the seven *co* mutants compared to wild type Landsberg *erecta* in LDs.
- Figure 2.4** The phenotypes of the seven *co* mutants, each carrying a different *co* mutant allele.
- Figure 2.5** The *co* mutant plants carrying the weak *co-4* allele and the strong *co-3* allele compared to wild type Landsberg *erecta*.
- Figure 2.6** The flowering times, represented as total leaf number, of wild type Landsberg *erecta* plants and *co-2* mutant plants grown under both LD and SD conditions.
- Figure 2.7** Examples of chimeric shoots.
- Figure 2.8** Examples of flowers with an altered number of petals and/or stamens.
- Figure 2.9** From the left, a wild type Landsberg *erecta* plant, a *co-4* heterozygote, a *co-4* homozygote, a *co-3* heterozygote, a *co-3* homozygote, a *co-7* heterozygote and a *co-7* homozygote.
- Figure 2.10** Bar charts showing the distribution of leaf numbers in the seven *co* mutant populations compared to the heterozygotes and wild type Landsberg *erecta*.
- Figure 2.11** Summary of leaf numbers of the seven homozygous *co* mutants in LDs compared to the heterozygotes and wild type Landsberg *erecta*.
- Figure 3.1** Southern blot of plant DNA extracted from wild type Landsberg *erecta* and the seven *co* mutants, digested with the enzymes indicated and probed using cosmid 129.
- Figure 3.2a** The sequencing gel that first identified the mutations in *co-4* and *co-6*.

- Figure 3.2b** The sequencing gel that first identified the mutations in *co-5* and *co-7*.
- Figure 3.3** The *CO* genomic region amplified by PCR and sequenced from all the *co* mutants.
- Figure 3.4** The DNA sequence of the *CO* ORF and the predicted amino acid sequence of the protein.
- Figure 3.5a** The predicted protein sequence of CONSTANS zinc finger 1 and zinc finger 2 with the prediction of secondary structure and the calculated probability of a secondary structure being formed underneath.
- Figure 3.5b** The predicted protein sequence of CONSTANS C-terminal basic domain with the prediction of secondary structure and the calculated probability of a secondary structure being formed underneath.
- Figure 3.6** The *CO* amino acid sequence with a number of features marked.
- Figure 3.7** Southern blot of RT-PCR products from wild type Landsberg *erecta* and the *co* mutants.
- Figure 3.8** Prettybox comparison of the predicted CONSTANS protein with CONSTANS-LIKE1 and CONSTANS-LIKE2.
- Figure 3.9** Prettybox comparison of the predicted protein sequence of the *CO* zinc fingers with a number of homologues from the GenBank database.
- Figure 3.10** Prettybox comparison of the predicted protein sequence of the CONSTANS zinc fingers with two proteins in the Swissprot database.
- Figure 3.11** Prettybox comparison of the predicted protein sequence of the *CO* C-terminal basic domain with a number of homologues in the GenBank database.
- Figure 3.12** A summary of some of the structural features of the *CO* protein thought to be important for its function.
- Figure 3.13** Sequence of the CONSTANS zinc finger region with secondary structure predicted by the PHD predict protein program.
- Figure 4.1** The strategy for generating new *co* alleles by gamma-irradiation of wild type pollen.
- Figure 4.2** Southern blot of plant DNA digested with *Hind* III and probed with the 2.3 kb *Eco*R I-*Pst* I fragment shown below.
- Figure 4.3** Southern blot of plant DNA digested with the restriction enzymes indicated and probed with the 950 bp probe shown below.
- Figure 4.4** The strategy used to isolate DNA adjacent to the *CO* promoter in *co-8*.

- Figure 4.5** Schematic diagram to illustrate how the DNA adjacent to the *CO* ORF in *co-8* was cloned by Inverse PCR.
- Figure 4.6** The sequence of the DNA which has been deleted from the *CO* locus in the *co-8* allele.
- Figure 4.7** Southern blot of PCR products using the primers indicated and DNA extracted from either wild type or the *co-8* mutant.
- Figure 4.8** DNA sequence of the junctions in the *co-8* allele.
- Figure 4.9** The PCR strategy used to establish whether an inversion or insertion had occurred in *co-8*.
- Figure 4.10** Genetic map positions of the YAC contigs at the *CO* locus and the contig at the second breakpoint in *co-8*.
- Figure 4.11** Southern blot of YAC DNA digested with *EcoR* I and *BamH* I.
- Figure 4.12** Southern blots of RT-PCR products probed with the *CO* cDNA, where amplification was with *CO* primers or with the *APETALA2* cDNA, where amplification was with the *AP2* primers.
- Figure 4.13** The *CO* amino acid sequence showing in red the amino acids which are predicted to be deleted from the ORF in *co-8*.
- Figure 4.14** The chromosomal rearrangement in *co-8*.
- Figure 4.15** The *co* mutants grown in LDs, as in Figure 2.4.
- Figure 4.16** The total leaf number of the eight *co* mutants compared to wild type Landsberg *erecta* in LDs.
- Figure 4.17** Days to flowering plotted against total leaf number for each *co* mutant and wild type grown in LDs.
- Figure 4.18** The *co-8* mutant compared to a *co-8* heterozygote and wild type Landsberg *erecta*.
- Figure 4.19** Summary of total leaf numbers of the *co* mutants grown in LDs, compared to the heterozygotes and wild type Landsberg *erecta*.
- Figure 5.1** Restriction maps of the constructs designed to express modified *CO* proteins.
- Figure 5.2** The sequence of the DNA fragment amplified to create the 35S:ZF construct.
- Figure 5.3** A wild type Landsberg *erecta* plant on the left, and a plant carrying the 35S:*co-7* construct on the right.

- Figure 5.4** Northern blot of RNA extracted from the 35S:*co-3*, 35S:ZF and 35S:*co-7* transformants.
- Figure 6.1** Prettybox comparison of the predicted protein sequence of the *Brassica napus* Bn1CON10 allele with *Arabidopsis* CO.
- Figure 6.2** Restriction map of the λ clone N.o.101 which contains the Bn1CON10 allele.
- Figure 6.3** From the left, a wild type Landsberg *erecta* plant, a *co-2 tt4* mutant and a *co-2 tt4* mutant carrying the *Brassica napus* CO homologue, Bn1CON10.
- Figure 6.4** Examples of phenotypes observed in some BnCO transformants.
- Figure 6.5** From the left, a wild type Landsberg *erecta* plant, a *co-2 tt4* mutant and two *co-2 tt4* mutants, each carrying the *Arabidopsis* promoter:*Brassica napus* ORF construct, ApBnCO, which did not complement the *co* mutant phenotype.
- Figure 6.6** Northern blots of RNA extracted from the BnCO transformants.
- Figure 6.7** Southern blots of RT-PCR products amplified from RNA extracted from the BnCO and ApBnCO transformants and untransformed control plants.

Tables

Table 1.1	Genes involved in the control of flowering time in <i>Arabidopsis thaliana</i> .
Table 1.2	Potential metal-binding residues in zinc finger-like proteins compared to CONSTANS.
Table 2.1	Flowering times and leaf numbers of the seven <i>co</i> mutants grown in LDs compared to wild type Landsberg <i>erecta</i> .
Table 2.2	Number of individuals with at least one flower with an altered number of petals and stamens, and the number of individuals with a chimeric inflorescence.
Table 2.3	Flowering times and leaf numbers in LDs of F ₁ progeny of crosses between each <i>co</i> mutant and Landsberg <i>erecta</i> , compared to the homozygous parental mutants and Landsberg <i>erecta</i> .
Table 3.1	The mutations in the seven <i>co</i> alleles, the domains of the CO protein they affect and the mean flowering times of the <i>co</i> mutants represented as total leaf numbers.
Table 3.2	The CO homologues with similarity over the zinc fingers.
Table 3.3	The CO homologues with similarity over the basic, C-terminal domain.
Table 4.1	The screen for lines which do not segregate early flowering plants in the M ₂ generation and which therefore may carry a new <i>co</i> allele.
Table 4.2	The analysis in the M ₃ generation of the six late flowering lines picked from the M ₂ screen and the identification of a line homozygous for wild type <i>LUTESCENS</i> and therefore potentially homozygous for a new <i>co</i> allele.
Table 4.3	Flowering times and leaf numbers of the eight <i>co</i> mutants grown in LDs compared to wild type Landsberg <i>erecta</i> .
Table 5.1	The kanamycin segregation analysis of the transformants in the T ₂ and T ₃ generations.
Table 5.2	Flowering time and leaf number analysis of the transformants in LDs and, where examined, SDs, compared to untransformed wild type Landsberg <i>erecta</i> and the <i>co-2</i> mutant.

Table 6.1	The kanamycin segregation analysis for the transformants in the T ₂ and T ₃ generations and the identification of families homozygous for the T-DNA/s to be used in further analysis of flowering time.
Table 6.2	Flowering time and leaf number analysis for the transformants in the <i>co-2 tt4</i> mutant background, in LDs and where applicable, SDs, compared to untransformed <i>co-2 tt4</i> and wild type Landsberg <i>erecta</i> .

Abbreviations

A, dATP	adenine, 2'-deoxyriboadenosine-5'-triphosphate
BAC	Bacterial Artificial Chromosome
bp	base pair
C, dCTP	cytosine, 2'-deoxycytidine-5'-triphosphate
°C	degrees Celsius
(CaMV) 35S	Cauliflower Mosaic Virus 35S RNA promoter
C-terminal	Carboxy terminal
cDNA	complementary DNA from transcribed RNA
CER	Controlled Environment Room
CHS	Chalcone Synthase
cm	centimetre
cM	centiMorgan
COPS	Controller of Phase Switching
dex	dexamethasone
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
DNP	Day Neutral Plant
EMS	ethyl methane sulphonic acid ester
ESSA	European Scientists Sequencing Arabidopsis
EST	Expressed Sequence Tags
$F_{1/2/3}$	F_1 = first filial generation; offspring resulting from first crossing of plants. F_2 , F_3 = progeny produced by self fertilization of F_1 , F_2 individuals.
FLIP	Floral Initiation Process
FT	Flowering Time
g	gram
G, dGTP	guanine, 2'-deoxyguanosine-5'-triphosphate
GA	Gibberellic Acid
GR	glucocorticoid receptor
h	hour(s)
IPCR	Inverse Polymerase Chain Reaction
K, Kan,	kanamycin
kb	kilobase
L	litre
LD(s)	long day(s)
LDP	Long Day Plant
Ler	Landsberg <i>erecta</i>
LG	Linkage Group
LN	Leaf Number
m	metre
M	Molar
$M_{1/2/3}$	M_1 = first generation after mutagenesis. M_2 , M_3 = progeny resulting from self-fertilization of M_1 , M_2 individuals.
Mb	Megabase
MBq	Mega Bequerels
mg	milligram
ml	millilitre
mm	millimetre
mM	milliMolar

mRNA	messenger RNA
μE	microEinstein
μg	microgram
μl	microlitre
μM	microMolar
ng	nanogram
NLS	Nuclear Localization Signal
nm	nanometre
N-terminal	Amino terminal
NTP	nucleoside triphosphate
OD	Optical Density
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
polyA	polyadenylation signal
QTL	Quantitative Trait Locus
RACE	Rapid Amplification of cDNA Ends
RFLP	Restriction Fragment Length Polymorphism
rif	rifampicin
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
RT-PCR	Reverse Transcription - Polymerase Chain Reaction
s	second
SD(s)	Short Day(s)
SDP	Short Day Plant
T, dTTP	thymine, 2'-deoxyribothymidine-5'-triphosphate
T-DNA	Transfer DNA
tet	tetracyclin
Tn	transformant
u	unit
UV-A	ultra-violet A
UV-B	ultra-violet B
var.	variety
WT	wild-type
w/v	weight per volume
YAC	Yeast Artificial Chromosome
5', 3'	5-prime, 3-prime

1 Introduction

1.1 Flowering

The transition from vegetative growth to flowering is one of the most dramatic developmental changes that occurs during the life cycle of flowering plants. Flowering enables the production of gametes, which in turn allow exchange of genetic information and ultimately the production of progeny. The survival of a species is dependent upon reproductive success, and it is strongly advantageous for plants, which are sessile, to recognize the most favourable environmental conditions in which to carry out reproductive development (Murfet, 1977). Many plants use environmental cues, particularly those which vary in predictable seasonal patterns, to regulate their transition to reproductive development. Hence the principal cues which regulate flowering are photoperiod (daylength), light quality and quantity, and temperature (Bernier *et al.*, 1981). The requirement for particular day lengths or for a period of low temperature (vernalization) to initiate flowering enables plants to become adapted to a particular location or latitude. This was first apparent in a variety of tobacco called Maryland Mammoth that required short days for flowering to occur. When grown in the northern United States this variety grew vegetatively throughout the summer months but flowered rapidly in the winter when the days became shorter (Garner and Allard, 1920). Similarly, many plants in northern latitudes flower only after an extended cold period, ensuring that flowering occurs in the spring and not during winter (Napp-Zinn, 1969). Other cues which are also important are water and nutrient availability, so that plants can survive times of drought or nutrient deprivation by flowering quickly and producing seeds before the parent plant dies.

Physiological and genetic studies have suggested a "multifactorial model" to explain how flowering is controlled (Bernier, 1988). This postulates that growth regulators and assimilates act as promoters and inhibitors of floral induction and these are required at a particular developmental stage and at an appropriate concentration for flowering to occur. The regulation of this process is proposed to be under both genetic and environmental control.

Plant species or varieties of the same species can be categorized according to their need for specific environmental conditions to initiate flowering. Those in which floral induction occurs at a predetermined time and proceeds at the same rate irrespective of environmental signals are known as autonomous species. Those which require absolute environmental conditions in order to flower are obligate or absolute species and plants where the transition to flowering is either hastened or delayed by environmental signals, but there is no absolute requirement for these signals, are facultative species (Bernier, 1988). *Arabidopsis thaliana* (*Arabidopsis*) belongs to the final category, with most ecotypes flowering eventually even under adverse conditions, but with environmental cues regulating the timing of the transition (Napp-Zinn, 1985). In most varieties of *Arabidopsis* long photoperiods and vernalization accelerate flowering.

1.1.1 Control of flowering by photoperiod

In the early 1920s Garner and Allard discovered that flowering and other developmental processes were regulated by daylength (Garner and Allard, 1920, 1923). They classified plants according to their flowering response to daylength, or photoperiod as they termed it. Three main categories emerged: short-day plants (SDP), which only flower (obligate species), or are accelerated to flower (facultative species), with fewer than a certain number of hours of light in each 24 h period; long-day plants (LDP) which only flower, or are accelerated to flower, with more than a certain number of hours of light in each 24 h

period, and day-neutral plants (DNP) which flower at the same time, irrespective of the daylength (autonomous species).

By exposing a variety of different plant species to light and dark cycles of different lengths it was concluded that SDP and LDP have different requirements for floral induction. For example, Hamner and Bonner (1938) showed that photoperiodic timekeeping in the SDP *Xanthium* is achieved by measuring the absolute duration of darkness. Flowering only occurred in this obligate species when the dark period exceeded 8.5 h. With sufficiently long nights, flowering was promoted even when coupled with a long light period. The importance of the absolute duration of darkness in SDPs is well documented (reviewed in Thomas and Vince-Prue, 1997), and it was originally thought that the measurement of nightlength was also critical in determining the flowering response of LDP. However, work on a number of LDP, for example, *Plantago lanceolata* (Snyder, 1948) suggested that for these plants a period of light longer than a critical length was necessary for the induction of flowering, rather than a period of darkness shorter than a critical length.

For many years it has been recognized that the main site for the perception of photoperiod is the leaf. Experiments by Knott (1934) with the LDP spinach first demonstrated this. Exposing only leaves to LDs promoted floral induction at the shoot apex, whereas exposing only the shoot apex to LDs resulted in plants which remained vegetative. Grafting experiments with a number of SDP and LDP suggested that leaves can independently produce a transmissible signal which is able to promote floral induction at the shoot apex (reviewed in Lang, 1952). This transmissible signal was assumed to be chemical, and named florigen by Chailakhayan in 1936. The model for the involvement of a specific floral hormone proposes that in the leaves, a hormone-like molecule is produced in inductive conditions (LDs for LDP, SDs for SDP, or as a consequence of other endogenous signals in DNP), and this molecule is transported to the apical meristem to induce flowering

(Chailakhayan, 1936). Substances that inhibit flowering have also been proposed, and in this case inductive conditions lead to the removal of this "anti-florigen". However, no single substance which can promote or prevent flowering in all species tested has been identified and the nature of florigen/anti-florigen is still unknown. More recent models (for example, Bernier, 1988) propose that floral induction occurs only when multiple signals are present at the shoot apex.

Grafting experiments with the LDP *Pisum* identified a graft-transmissible mobile floral stimulus and a mobile floral inhibitor that are produced in the leaves. Genes controlling the production of each have been identified from studying mutants which are affected in floral initiation (Weller *et al.*, 1997). The *gigas* (*gi*) mutant, for example, which is late flowering in LDs and SDs, appears to be deficient in a mobile floral stimulus (Beveridge and Murfet, 1996). In grafting experiments, wild type donor plants were able to promote flowering in mutant *gi* shoots, suggesting that a stimulus produced by the wild type plants was able to move across the graft union, and the *gi* mutant could respond normally to this stimulus. The *gi* mutation, therefore, is proposed to block the synthesis of the floral stimulus. The identity of the *Gi* gene, however, and the chemical nature of the floral stimulus are still unknown.

1.2 *Arabidopsis thaliana*

Arabidopsis thaliana, a member of the Cruciferae, or Brassicaceae, was first described by the physician Johannes Thal, in a German book describing the flora of the Hartz Mountains, published in 1577.

"..at the top, tiny white florets bloom and then drop to be followed by small, thin, oblong seed pods....The pods are packed with tiny, reddish seeds. The plant is widely observed on the hillsides at the beginning of spring..."
(Johannes Thal, 1577; transl. Ingrid Walton).

At the time it was known as *Pilosella siliquata minor* but from 1842, in honour of Johannes Thal, it became known as *Arabidopsis thaliana*. The species is distributed mostly in the temperate zones of Europe, West and Central Asia, Japan, North, East and South Africa, Australia and North America, with its centre of distribution generally considered to be Western Europe (Redei, 1969).

In 1907 the haploid chromosome number of *Arabidopsis thaliana* was established as five by Friedrich Laibach (Laibach, 1907). In 1943 he published a paper which described the advantages of using *Arabidopsis thaliana* as an experimental organism (Laibach, 1943), and this paper laid the foundations for the next five decades of research.

Arabidopsis was not a species of choice for physiological analyses of flowering because of its facultative response to environmental conditions, and because its small size prevented the use of approaches such as grafting and harvesting of large amounts of tissue for physiological analysis. However it has become the preferred model for genetic analysis due to the high level of natural variation in flowering time (Rédei *et al.*, 1974), the ease with which mutants can be isolated (Hussein, 1968) and its utility for molecular analysis (Meyerowitz, 1987).

A number of features make *Arabidopsis* a convenient model species for classical and molecular genetics. *Arabidopsis* is self-fertile, and can also be cross-fertilized relatively easily. It has a short generation time of six to eight weeks and a high seed yield, producing up to 10,000 seeds per plant. Its small size and minimal growth requirements mean that large populations can be maintained in greenhouses or sterile tissue culture facilities. Mutagenesis screens can be performed with large numbers of seeds because of the small seed size, and large M₂ populations can be grown in a relatively small area to look for specific mutations. Chemical mutagens, such as ethyl methane sulphonate (EMS) in addition to fast neutron, X- and γ -irradiation have been used successfully in mutagenesis experiments (reviewed in Lightner and Caspar, 1998, Feldman *et al.*, 1994 and Rédei and

Koncz; 1992) The isolation of a large number of visible mutations allowed the construction of a genetic map of the *Arabidopsis* genome and the production of multiply marked lines on which to map new mutations (Koornneef *et al.*, 1983; Koornneef, 1990). *Arabidopsis* has a number of features that facilitate the cloning of genes like *CO*, of which nothing is known except the phenotype caused by mutations within the gene. The genome of *Arabidopsis* is considerably smaller than most other higher plants. Current estimates put the size of the haploid genome at approximately 100 Mb (Leutwiler *et al.*, 1984; Arumuganathan and Earle, 1991) present on five small chromosomes, containing relatively little repetitive DNA (Leutwiler *et al.*, 1984). These factors have facilitated a number of gene cloning methods. T-DNA tagging, where foreign DNA is integrated into the genome by *Agrobacterium*-mediated transformation, benefits from the high proportion of the *Arabidopsis* genome that is made up of genes (Bevan *et al.*, 1998; Feldman *et al.*, 1991) and from the ease by which *Arabidopsis* can be transformed (Bechtold *et al.*, 1993). Genomic subtraction, in which the DNA lost in deletion alleles can be isolated (e.g. the *gal-3* allele; Sun *et al.*, 1992) also benefits as the method relies on the hybridization of mutant and wild type DNA, a factor which is affected by the complexity of the genome (Wetmur and Davidson, 1968). However, both these methods rely on the availability of either T-DNA-tagged alleles or known deletion alleles. Consequently, map-based cloning by chromosome walking would be the preferred method in the absence of such alleles. This method is also greatly facilitated in *Arabidopsis* by the small genome size and particularly the relatively small amount of repetitive DNA, which would otherwise interrupt a chromosome walk. Map-based cloning involves fine-mapping the gene of interest to a small region between two RFLP markers and isolating cloned genomic DNA in the region to build a physical map. The DNA isolated can be subcloned into small fragments for complementation of the mutant phenotype by *Agrobacterium*-mediated transformation. In *Arabidopsis* the existence of detailed RFLP and genetic maps (Koornneef *et al.*, 1983;

Chang *et al.*, 1988; Nam *et al.*, 1989), the generation of a number of *Arabidopsis* genomic Yeast Artificial Chromosome libraries (Ward and Jen, 1990; Grill and Somerville, 1991; Guzman and Ecker, 1988; Ecker, 1990) and the ordering of 20,000 cosmid clones into contigs (Hauge *et al.*, 1991; Hauge and Goodman, 1992) facilitated the first chromosome walks. More recently, map-based cloning has been aided by the generation of YAC and BAC (Bacterial Artificial Chromosome) libraries with larger inserts (Creusot *et al.*, 1995), the availability of many more RFLP markers and the development of PCR markers (Bell and Ecker, 1994; Konieczny and Ausubel, 1993; Reiter *et al.*, 1992) and the creation of YAC contig maps with the aim of producing a complete physical map of the *Arabidopsis* genome. Physical maps of chromosomes 2 (Zachgo *et al.*, 1996), 4 (Schmidt *et al.*, 1995) and 5 (Schmidt *et al.*, 1997) have been published and these have provided a basis for sequencing the *Arabidopsis* genome. An international collaboration between *Arabidopsis* research groups, the *Arabidopsis thaliana* Genome Research Project, aims to completely sequence the *Arabidopsis* genome by the year 2004. The analysis of 1.9 Mb of contiguous sequence from chromosome 4 was recently published by the EU Arabidopsis Genome Project (Bevan *et al.*, 1998). Extensive physical maps for all five chromosomes are now available via the Internet (*Arabidopsis thaliana* Database (AtDB) URL:<http://genome-www.stanford.edu/>). In addition, the sequence of transcripts is being determined (expressed sequence tags (ESTs) URL:<http://www.tigr.org/> Höfte *et al.*, 1993; Newman *et al.*, 1994) and genomic sequence information is becoming available via the internet at a rate of approximately 1 Mb per month (I. Bancroft, personal communication).

In 1992 the first reports of genes isolated from *Arabidopsis* by map-based cloning were published. The *ABI3* gene, involved in the response to the hormone abscisic acid, was mapped to position 23.5 cM on chromosome 3 (Koornneef, 1987) and after further fine mapping was located to a cosmid in an existing contig (Hauge *et al.*, 1991). This cosmid

was used successfully to complement the *abi3* mutant phenotype (Giraudat et al., 1992). To isolate the *fad3* (*fatty acid desaturase3*) gene, a contig was built over the *fad3* locus by using closely linked RFLP markers as probes against YAC libraries. These YACs were used to isolate a cDNA for a highly expressed mRNA specific to seeds from a *Brassica napus* cDNA library (Arondel et al., 1992) and this was shown to complement the *Arabidopsis fad3* mutant phenotype.

The ability to regenerate whole plants from a number of *Arabidopsis* tissues, and the relatively small amount of space required for the sterile culturing of *Arabidopsis* plants lead to the development of direct DNA uptake and *Agrobacterium tumefaciens*-mediated transformation methods (Damm et al., 1989; Valvekens et al., 1988). These methods have been largely superseded by *in vivo* transformation methods. For example, submerging plants in the desired *Agrobacterium* strain while under vacuum enables transformation of cells that give rise to gametes and enables the selection of transformants among the progeny of vacuum-infiltrated plants (Bechtold et al., 1993). This method facilitates the generation of large numbers of transformants with minimum effort.

1.2.1 Flowering in *Arabidopsis thaliana*

Arabidopsis is an annual species in which the vegetative and reproductive developmental phases are temporally separated. It belongs to a group of plants characterized by the formation of a rosette during vegetative development, as a consequence of reiterative leaf formation in the absence of stem elongation. The first pair of leaves, arising from the apical meristem, develop opposite one another in a decussate phyllotaxy. Subsequent leaves are produced 104° apart in a spiral phyllotaxy, forming the rosette. At a time determined by environmental conditions and the genotype of the individual the vegetative shoot apical meristem becomes an inflorescence meristem, and starts to produce flower buds instead of leaves. This transition is immediately followed by the elongation of the internodes between

the last few leaves and between the flowers, giving rise to a stem carrying cauline leaves (to distinguish them from the rosette leaves) and flowers, still in the characteristic spiral phyllotaxy. Subsequently, indeterminate inflorescence meristems develop in the axils of the cauline leaves and some of the rosette leaves, later to develop as co-florescences with a structure similar to the primary inflorescence (main flowering stem). The shoot apical meristem remains indeterminate, continuing to form flowers indefinitely until the plant starts to senesce (reviewed in Madueño *et al.*, 1996).

Arabidopsis flowers are composed of four whorls each consisting of a distinct organ type. There are four sepals in the first whorl. These are the first floral organs to differentiate, and they grow to overlie the developing floral primordium. Subsequently four petals in the second whorl and six stamens in the third whorl develop, and finally two fused carpels arise in the fourth whorl, at the centre of the floral meristem.

The transition from vegetative to reproductive development is marked by the establishment of floral fate in the apical meristem and the suppression of leaf production (Hempel and Feldman, 1994). Genes known as floral meristem identity genes are required for primordia to develop as flowers rather than leaves (Schultz and Haughn, 1993). A number of these have been identified, including *LEAFY* (*LFY*), *APETALA1*, *APETALA2* (*API*, *AP2*) and *CAULIFLOWER* (*CAL*) (Haughn *et al.*, 1995). Mutations in these genes cause the meristems to produce inflorescence-like structures rather than flowers, with the most severe effect seen in *lfy* mutants (Schultz and Haughn, 1991; Weigel *et al.*, 1992). Ectopic expression of the wild type *LFY* or *API* genes from the CaMV 35S promoter converted shoots into flowers, indicating that these genes are sufficient to confer floral identity on the meristem. However, before flowering these transformants underwent a period of vegetative growth, indicating that additional factors are required at the apical meristem for it to be

able to respond to the expression of the floral meristem identity genes and initiate the flowering process (Weigel and Nilsson, 1995; Mandel and Yanofsky, 1995).

In addition to genes which confer floral identity on developing primordia, genes which are required for the maintenance of the vegetative state or of the indeterminacy of the inflorescence meristem have been identified. Mutations in the *TERMINAL FLOWER* (*TFL*) gene cause the shoot apical meristem to become determinate, converting it into a floral meristem which terminates in a flower (Shannon and Meeks-Wagner, 1991; Alvarez *et al.*, 1992). This is similar to the phenotype caused by ectopic *LFY* expression, and double mutant analysis indicated that *LFY* and *TFL* have antagonistic roles (Shannon and Meeks-Wagner, 1993). Mutations in *EMBRYONIC FLOWER1* and 2 (*EMF1*, *EMF2*) cause the shoot apical meristem to be converted to an inflorescence meristem on germination (Sung *et al.*, 1992). The *EMF* genes, then, appear to negatively regulate the transition from vegetative to reproductive development, and this has led to the suggestion that they encode floral repressors whose activity decreases during development, allowing the switch to floral development to occur (Chen *et al.*, 1997). A model for this process has been suggested, where the shoot apical meristem matures and the floral repressor, the Controller of Phase Switching (COPS), decreases until it is below a critical threshold at which point the Floral Initiation Process (FLIP) is activated (Schultz and Haughn, 1993).

Regulation of the timing of inflorescence and flower development therefore requires the control of the expression and activity of the floral meristem identity genes. The expression of meristem identity genes is regulated by environmental signals, such as daylength (Blasquez *et al.*, 1997) and this is likely to occur via the action of flowering time genes (Simon *et al.*, 1996).

1.2.1.1 Environmental control of flowering in *Arabidopsis*

Arabidopsis is a facultative long day plant (LDP) which shows no absolute requirement for any specific environment but will eventually flower even in unfavourable conditions (Rédei *et al.*, 1974). However, a number of factors can hasten or delay the transition from vegetative to reproductive growth. Long photoperiods, cold treatment (vernalization) and particular wavelengths of light greatly accelerate the onset of flowering (Martínez-Zapater *et al.*, 1994).

1.2.1.1.1 Photoperiod and light quality

Light plays an important regulatory role in floral induction, although in *Arabidopsis* it is not essential for flowering. Wild type plants grown in the dark in liquid sucrose media were able to flower at the same time as plants grown in continuous light (Araki and Komeda, 1993a). Under these conditions the late flowering *co-1* and *gi-2* mutants (Koornneef *et al.*, 1991) flowered as early as wild type, suggesting that the products of the genes are not required for flowering in darkness.

Flowering of *Arabidopsis* responds to light in two ways; by flowering earlier under long than short photoperiods, and by being influenced by light quality (i.e. exposure to different wavelengths of light). Plants detect light through a number of photoreceptors: the phytochromes, which detect red and far-red light, the cryptochromes, which detect blue and UV-A light and the currently undiscovered UV-B receptors (Chory, 1993). Flowering in *Arabidopsis* is accelerated by far-red light and blue light, and delayed by red light (Eskins, 1992), suggesting that both phytochrome and cryptochrome are involved in controlling flowering time.

A number of *Arabidopsis* mutants which were originally identified because they are defective in light perception or signal transduction also show altered flowering time. These mutants are affected in photomorphogenesis, developing an elongated hypocotyl in the

light, a phenotype normally associated with dark grown etiolated plants. Mutant screens in different wavelengths of light (white, blue, far-red) have identified mutations in genes encoding different light receptors or signal transduction components. The *hy1*, *hy2* and *hy6* mutants (*elongated hypocotyl*; Koornneef *et al.*, 1980) are deficient in the synthesis or attachment of the chromophore, which is essential for the function of all phytochromes (Parks and Quail, 1991). These mutants are early flowering in both LDs and SDs. Mutants carrying lesions in the individual structural genes encoding the photoreceptor proteins have also been isolated, and these mutants are more useful when assigning roles to the different phytochrome or cryptochrome species. Mutations in the gene encoding the light-labile phytochrome A (PHYA, or type I phytochrome) cause late flowering in LDs with low fluence rate daylength extensions (Johnson *et al.*, 1994). The mutants also show an elongated hypocotyl in far-red light, suggesting that PHYA plays a role in detecting low light and light with low red:far-red (R:FR) ratios (Whitelam *et al.*, 1993) and promotes flowering in response to these conditions. This is consistent with the observation that a far-red extension to short days can induce flowering (Mozley and Thomas, 1995).

Red light, on the other hand, has been shown to inhibit flowering in *Arabidopsis* (Eskins, 1992). Mutations in the gene encoding the light-stable phytochrome B (PHYB, type II phytochrome), which is proposed to be responsible for the detection and response to red light (Quail *et al.*, 1995) result in early flowering in both LDs and SDs (Halliday *et al.*, 1994). The inhibition of flowering by red light, therefore, is mediated by PHYB, and possibly other light-stable forms of phytochrome (Halliday *et al.*, 1994). The phenotype of the *phyB* mutant, *hy3*, mimicks the shade avoidance response, where plants shaded by neighbouring vegetation flower early and have elongated stems (Reed *et al.*, 1993). In shade, the ratio of R:FR light decreases, because FR light is reflected from leaves and R light is absorbed (Smith *et al.*, 1990). In these conditions, PHYB is in the inactive P_r form, removing the inhibition of flowering and allowing the plants to flower. In the open, where

the R:FR ratio is higher, PHYB is in the active P_{fr} form, repressing the shade avoidance response and inhibiting flowering (Yanofsky *et al.*, 1995).

Photobiological studies have clearly demonstrated that flowering in *Arabidopsis* is promoted by blue light (Eskins, 1992). Phytochrome can be weakly activated by blue light, in addition to its response to red and far-red light, suggesting that the blue light response could also be mediated through this photoreceptor (Bernier *et al.*, 1993). However, the existence of specific blue light receptors has recently been demonstrated. The *hy4* mutant, which is defective in blue light-inhibition of hypocotyl elongation, was identified by Koornneef *et al.* (1980). The *HY4* gene encodes a protein with homology to a microbial flavoprotein that catalyses blue light-dependent reactions (Ahmed and Cashmore, 1993). This gene is now known as *CRYPTOCHROME1* (*CRY1*; Lin *et al.*, 1996). A number of mutants with reduced blue light response have been found to be allelic to *hy4* and their flowering time has been analyzed. Although *hy4* itself does not show a significantly altered flowering time phenotype, other *hy4* alleles isolated in the Columbia ecotype show a variety of responses to daylength, where one has a wild type phenotype in LDs and is early in SDs (*hy4-105*) and others are delayed under both LDs and SDs (*hy4-101*, *-102*, *-103*) (Zagotta *et al.*, 1996; King and Bagnall, 1996). The late flowering phenotype of *hy4* alleles in SDs was corrected to that of wild type by red and far-red night breaks, indicating that cryptochrome is the receptor which is responsible for the blue light-promotion of flowering in *Arabidopsis* (King and Bagnall, 1996). The *hy4* mutations have little effect on flowering time in LDs, suggesting that the *CRY1* receptor is responsible for a specific blue light response, promoting flowering in SDs or in response to certain light regimes. That blue light plays an important role in photoperiodic induction of flowering in LDs has recently been demonstrated by the discovery that the *FHA* gene encodes a second blue light receptor, previously known as *CRY2*, identified by homology to *CRY1* (Lin *et al.*, 1996; Guo *et al.*, 1998). The *FHA* gene has been proposed to act in a pathway that promotes

flowering in response to LDs, because mutations at this locus delay flowering in LDs but not SDs (Koornneef *et al.*, 1991; section 1.2.1.3.1). The behaviour of the *cry2* mutants in a number of different light qualities suggested that CRY2 and PHYB operate antagonistically to regulate flowering time, and that the role in LDs of PHYB is to inhibit flowering, and the role of CRY2 is to repress PHYB function (Guo *et al.*, 1998). The interaction of *FHA* and other flowering time genes is discussed later.

Arabidopsis is a facultative LDP, flowering eventually in all conditions but accelerated by exposure to long photoperiods. This acceleration is seen most dramatically in early flowering ecotypes, such as Landsberg *erecta*, which flowers in 2-3 weeks with 6-7 leaves in a 16 h photoperiod, and after 6 weeks with more than 30 leaves in short days of less than 10 h. The ability to respond to LDs, however, depends on the age of the plant. Experiments in which plants were moved from non-inductive SDs into LDs demonstrated that seedlings which are 4 days old or less are not responsive to long days. Seven day old seedlings required 5 long days to promote flowering, while 20 day old plants needed just 1 long day (Mozley and Thomas, 1995; Corbesier *et al.*, 1996). Therefore, as *Arabidopsis* plants mature they become more responsive to long day induction.

The measurement of daylength is thought to be facilitated by an endogenous circadian clock (Thomas and Vince-Prue, 1997). This is entrained to a period of approximately 24 h, and is reset by signals from light receptors (Millar *et al.*, 1995). A number of processes in plants are controlled by the circadian clock including leaf movements and expression of many genes. In *Arabidopsis*, for example, the transcription of the light regulated genes *chlorophyll A/B binding protein 2* (*CAB2*; Millar and Kay, 1991) and the gene encoding the small subunit of rubisco (*Ribulose Bisphosphate Carboxylase*, *RBCS*) is controlled by the clock. Rhythmic leaf movements and transcription of reporter genes under the control of the promoters of these circadian-regulated genes has provided a means to study the clock

and the environmental signals which interact with it. These experiments have demonstrated that in *Arabidopsis* the period of the clock, the time taken for one cycle, is shorter in light than in dark, and that this response is mediated by blue light photoreceptors and phytochrome (Millar *et al.*, 1995). A number of mutants which affect the circadian clock and also have an altered flowering time have recently been reported. The most well characterized is the *elf3* mutant (*early flowering*) which shows photoperiod-insensitive early flowering and lacks rhythmicity in circadian-regulated processes in continuous light (Hicks *et al.*, 1996a; Zagotta *et al.*, 1996). However, the mutant retains rhythmicity in continuous dark, suggesting that the *ELF3* gene plays a role in the signalling of light to the clock rather than being a component of the clock mechanism (Hicks *et al.*, 1996a). The mutant also has an elongated hypocotyl in blue light which might indicate that blue light, perhaps detected via CRY2, is involved in signalling to the clock and allowing the plant to distinguish between LDs and SDs (Suárez-López and Coupland, 1998). Other mutants showing altered flowering time and disrupted circadian rhythms are the late flowering *lhy* mutant (Coupland, 1997) which is arrhythmic (Schaffer *et al.*, 1998) and the early flowering *det1* and *esd4* mutants which have a short period compared to wild type (Millar *et al.*, 1995; Carre, Murtas, Reeves and Coupland, unpublished results). The *LHY* gene has been proposed to act, like *FHA*, in the long day promotion pathway (Coupland, 1997) and might suggest a role for the circadian clock in that pathway.

1.2.1.1.2 Vernalization

For many plant species a period of cold treatment is absolutely required for, or accelerates, floral induction. The process of exposing a plant to low temperatures in order to induce or accelerate flowering is known as vernalization. Optimum temperatures for induction or acceleration of flowering range between 1° and 7°C, for 1 to 3 months, which is similar to a winter season in many temperate zones (Dennis *et al.*, 1996). Physiological experiments

with celery established that the site of detection of cold treatment is the apical meristem (Curtis and Chang, 1930). However, studies on a number of different tissues with mitotically dividing cells from a range of species suggest that dividing cells in any part of a plant can respond to vernalization (Dennis *et al.*, 1996). For example, in *Arabidopsis* root explants subjected to a period of cold treatment developed into flowering shoots (Burn *et al.*, 1993a). *Arabidopsis* has a number of late flowering winter annual ecotypes which respond strongly to vernalization by flowering early, and other ecotypes which are less responsive (Bagnall, 1993).

There are features of the vernalization response that suggest it might occur through an epigenetic mechanism (Dennis *et al.*, 1996). It has been proposed that vernalization causes general DNA demethylation, and that the promotion of flowering is caused by the demethylation of promoters of genes critical for floral induction (Burn *et al.*, 1993a). One of these could be the gene encoding kaurenoic acid hydroxylase which is an important gene in gibberellic acid biosynthesis. Gibberellic acid levels were shown to increase in vernalized radish plants (Suge, 1970) leading to the original proposal that GA mediated the vernalization response. However, more recent work with *Arabidopsis* mutants, which are insensitive to vernalization but still flower early on application of exogenous GA, indicate that while GA is important for the promotion of flowering, vernalization occurs by a different mechanism (Chandler and Dean, 1994).

A number of late flowering *Arabidopsis* mutants which can be restored to early flowering by vernalization have been characterized (Koornneef *et al.*, 1991). These mutants are proposed to be defective in the autonomous flowering pathway and are still responsive to environmental stimuli (see later). These mutants formed the basis of a genetic approach to study the vernalization response. Vernalization-insensitive mutants were isolated after mutagenesis of the late flowering, vernalization responsive *fca* mutant (Chandler *et al.*, 1996). The two mutants isolated, *vrn1* and *vrn2*, are late flowering in all environmental

conditions because of the presence of the *fca* mutation, but cannot be restored to early flowering by vernalization. In the absence of the *fca* mutation the *vrn* mutants do not delay flowering, indicating that they are not classical late flowering mutants but mutants in the vernalization response (Levy and Dean, 1998). The cloning and characterization of *VRN1* and *VRN2* should help in the identification of the processes important in the promotion of flowering by vernalization.

1.2.1.2 Genetic analysis of the control of flowering time

Genetic analysis of the control of flowering time in *Arabidopsis* was facilitated by the availability of natural populations displaying variation in their flowering times, and by the identification of flowering time mutants after mutagenesis. Studying the effect of monogenic mutations at different loci in the same genetic background made it possible to study the interactions between mutations and to define genetic pathways involved in the transition to flowering. However, the analysis of natural variation in flowering time, expressed as differences between ecotypes, revealed loci not initially identified by mutagenesis of only a few ecotypes. This natural variation is clearly important in contributing to the control of flowering time in natural environments.

1.2.1.2.1 Natural variation in flowering time

Extensive variation in flowering time exists among *Arabidopsis* natural populations (ecotypes) (Napp-Zin, 1985). This variation has been studied since the early 1950s by comparing and crossing selected inbred lines and analyzing the segregation of the loci involved (Laibach, 1951). To illustrate the extent of variation found, 32 different ecotypes from a variety of locations in the Northern hemisphere were grown in a number of different daylength environments with and without vernalization, and the effect of these environments on flowering time recorded (Karlsson *et al.*, 1993). Flowering time in long

days ranged from Landsberg *erecta* which flowered with an average of 4.2 leaves to the late flowering TS-2 ecotype from Spain which flowered with around 30 leaves.

Genetic analysis of the progeny of a cross between the late flowering ecotype Stockholm and the early flowering Limberg-5 produced genotypes with very different flowering times caused by single gene differences. Analysis of different segregants established that several genes were responsible for the flowering time variation between the progenitors (Napp-Zinn, 1969). The major effect on flowering time was conferred by the *FRIGIDA* (*FRI*) locus, dominant alleles of which were responsible for the late flowering of Stockholm. *FRI* was found to be epistatic to the other genes involved, including *KRYOPHILA* (*KRY*) and *JUVENALIS* (*JUV*). *FRI*, which maps to the top arm of chromosome 4, is probably also responsible for the late flowering of a number of other late ecotypes including Piztal and Innsbruck (Burn *et al.*, 1993b). The locus *FLOWERING ALTERED* (*FLA*), which was responsible for the late flowering of ecotypes Saint Feliu (SF-2) and Leiden (Le-0) is probably also allelic to *FRI* (Lee *et al.*, 1993; Clarke and Dean, 1994).

The late flowering phenotype conferred by dominant alleles of *FRI* is suppressed in the early flowering Landsberg *erecta* and C24 ecotypes (Lee *et al.*, 1994a; Koornneef *et al.*, 1994). Dominant alleles at a second locus, *FLOWERING LOCUS C* (*FLC*), on the top arm of chromosome 5, are required for the late flowering phenotype of *FRI* to be observed. Landsberg *erecta* and C24 have early *FLC* alleles which suppress the late flowering phenotype of *FRI* and mutations such as *ld* (Lee *et al.*, 1994a) and *fld* (Sanda and Amasino, 1996a), which were originally isolated in the Columbia background. Landsberg *erecta* alleles of these two genes were not isolated during extensive mutagenesis experiments to identify late flowering mutants (Koornneef *et al.*, 1991). *FLC* is considered to be a dominant inhibitor of flowering, which is itself inhibited by *LD*, *FLD* and possibly other genes in the autonomous flowering pathway (Aukerman and Amasino, 1996).

Analysis of the very late flowering ecotype, Skye, has identified a third locus responsible for natural flowering time variation. This locus, *AERIAL ROSETTE* (*ART*) is on chromosome 5. A dominant *ART* allele from Skye, in combination with *FRI*, gives rise to plants with aerial rosettes which develop from axillary meristems that would normally yield inflorescences (Grbic and Bleecker, 1996; Grbic and Gray, 1997).

In addition to these dominant alleles with large effects there appear to be many other natural alleles with smaller effects, which have been identified by Quantitative Trait Locus (QTL) analysis. Twelve QTLs affecting flowering time were found between the Landsberg *erecta* and Columbia ecotypes (Jansen *et al.*, 1995), some of which map close to known flowering time mutations.

1.2.1.2.2 Mutations affecting flowering time

The isolation of induced mutations affecting flowering time of *Arabidopsis* began in the early 1960s (McKelvie, 1962; Rédei, 1962). Most mutagenesis screens were performed using early flowering ecotypes in order to isolate late flowering mutants, which are relatively easy to screen for in standard long day conditions. Recently, however, a number of early flowering mutants have been isolated by screening mutagenized populations in non-inductive short day conditions or in tissue culture.

In addition to mutants which only affect the transition to flowering, there are a number of other mutants which were first identified on the basis of other phenotypes, and subsequently found to also display late or early flowering. Many mutants affect the rate of development of the plant by, for example, retarding growth. These mutants were excluded from the analysis of late flowering mutants by determining the rate of leaf initiation and studying those mutants that form leaves at the same rate as wild type but have more or fewer leaves than wild type at the time of flowering (Koornneef *et al.*, 1991).

1.2.1.2.2.1 Late-flowering mutants

In 1962, two papers which described the identification of late flowering mutants from mutagenesis screens were published. McKelvie (1962) identified 700 mutations at 190 different loci after radiation and chemical mutagenesis of the ecotypes Estland and Limberg. Six of these mutants were late flowering and were named *F* (*Florens*), *f1*, *f2*, *f3*, *f4* and *f5*. The *F* mutant, which was dominant and responded strongly to vernalization, was subsequently found to be allelic to *FRI*, although it is now unclear whether this was a mutant or a contaminating natural variant at this locus (Koornneef *et al.*, 1994).

Rédei (1962) identified three very late flowering mutants with no other obvious pleiotropic phenotypes, in the early ecotype Columbia. These were named *constans*, *gigantea* and *luminidependens* (*co*, *gi* and *ld*, respectively) and were the first of twelve mutants which were subsequently studied in detail. This thesis is concerned with one of these mutants: *constans*.

In 1968 Hussein isolated a number of other late flowering mutants, *fb*, *fca*, *fe*, *fg*, *fve* and *fy* from the early ecotype Landsberg *erecta*. These were located on the genetic linkage map by Koornneef *et al.* (1983). In 1991 Koornneef *et al.* described allelism tests with 42 independently isolated mutations causing late flowering, including mutations at the six loci identified by Hussein, and these were found to correspond to 11 loci. In addition, these tests showed that *fg* and *fb*, originally isolated by Hussein, were allelic to Rédei's *co* and *gi* mutations, respectively. None of the 42 mutations were allelic to Rédei's *ld* mutant although subsequent alleles were found in ecotype Wassilewskija (*Ws*; Lee *et al.*, 1994a). Later it became clear that the late flowering phenotype caused by *ld*, originally isolated in Columbia, is suppressed by the Landsberg *erecta* allele at the *FLC* locus (Lee *et al.*, 1994a; Koornneef *et al.*, 1994).

The twelve genes which correspond to the most studied late flowering mutants are *LD*, *CO*, *GI*, *FE*, *FT*, *FD*, *FY*, *FCA*, *FHA*, *FPA*, *FVE* and *FWA* (Koornneef *et al.*, 1991; Table 1.1). A number of other late flowering mutants have been isolated. For example, in the Columbia background, which contains the dominant *FLC* allele, the *fld* mutant was isolated. The *fld* mutant phenotype, like that of *ld*, is suppressed in the early Landsberg *erecta* background (Sanda and Amasino, 1996a).

Many late flowering mutants have been identified on the basis of other phenotypes or because they are late flowering only under certain environmental conditions. Mutants affected in the processes regulating light perception or transduction, for example *hy4* and *phyA* (Ahmad and Cashmore, 1993; Whitelam and Harberd, 1994), or in hormone metabolism, *gal* and *gai* (Wilson *et al.*, 1992) can also exhibit late flowering. A mutation that disrupts circadian clock function, *lhy*, also causes a late flowering phenotype (Simon and Coupland, 1996). Carbohydrate metabolism mutants are often late flowering, as is the case for *cam1* (*carbohydrate accumulation mutant1*), which is late flowering under LD conditions, as well as *sex1* (*starch excess1*) and *pgm1* (*phosphoglucosmutase1*) which are both late flowering under SD conditions (Eimert *et al.*, 1995). The late flowering *gigantea* (*gi*) mutant (Rédei 1962; Koornneef *et al.*, 1991) was also found to have increased starch levels in the leaves (Araki and Komeda, 1993b; Eimert *et al.*, 1995). *vrn1* and *vrn2*, isolated in an *fca* background, are late flowering after vernalization treatment, which normally corrects the late flowering phenotype of the *fca* single mutant (Chandler and Dean, 1994). All late flowering mutants described here, and a number of additional ones which are only late under certain conditions, are summarized in Table 1.1.

Genes involved in the control of flowering time	Flowering response of mutant compared to wild type	Reference
Daylength detection		
<i>CO</i> (<i>CONSTANS</i> = <i>FG</i>)	Late LD	1
<i>GI</i> (<i>GIGANTEA</i> = <i>FB</i>)	Late LD	2
<i>FD</i>	Late LD, SD	2
<i>FE</i>	Late LD, SD	2
<i>FHA</i> (= <i>CRY2</i>)	Late LD	3
<i>FT</i>	Late LD, SD	9
<i>FWA</i>	Late LD, SD	2
Endogenous flowering response		
<i>ART</i> (<i>AERIAL ROSETTE</i>)	Late LD, SD	4
<i>FCA</i>	Late LD, SD	5
<i>FLC</i> (<i>FLOWERING LOCUS C</i>)	Late LD, SD	6
<i>FLD</i> (<i>FLOWERING LOCUS D</i>)	Late LD, SD	7
<i>FPA</i>	Late LD, SD	2
<i>FRI</i> (<i>FRIGIDA</i> = <i>FLA</i>)	Late LD, SD	8
<i>FVE</i>	Late LD, SD	2
<i>FY</i>	Late LD, SD	2
<i>LD</i> (<i>LUMINIDEPENDENS</i>)	Late LD, SD	10
<i>SIN</i> (<i>SHORT INTEGUMENT</i>)	Late LD, SD	11
Light detection		
<i>COP1</i> (<i>CONSTITUTIVE PHOTOMORPHOGENIC 1</i>)	Early SD	12
<i>DET1</i> (<i>DEETIOLATED 1</i>)	Early SD	13
<i>ELF3</i> (<i>EARLY FLOWERING 3</i>)	Early LD, SD	14
<i>HY1</i> (<i>LONG HYPOCOTYL 1</i>)	Early LD, SD	15
<i>HY2</i> (<i>LONG HYPOCOTYL 2</i>)	Early LD, SD	15
<i>HY3</i> (<i>LONG HYPOCOTYL 3</i> = <i>PHYB</i>)	Early LD, SD	16
<i>HY4</i> (<i>LONG HYPOCOTYL 4</i> = <i>CRY1</i>)	Late SD	17
<i>HY6</i> (<i>LONG HYPOCOTYL 6</i>)	Early LD, SD	18
<i>PEF1</i> (<i>PHYTOCHROME SIGNALLING EARLY FLOWERING 1</i>)	Early LD, SD	19
<i>PEF2</i> (<i>PHYTOCHROME SIGNALLING EARLY FLOWERING 2</i>)	Early LD, SD	19
<i>PEF3</i> (<i>PHYTOCHROME SIGNALLING EARLY FLOWERING 3</i>)	Early LD, SD	19

<i>PHYA</i> (<i>PHYTOCHROME A</i>)	Late in low fluence day extensions (SD + extension with low fluence light)	18
<i>TED1</i> (<i>SUPPRESSOR OF DET 1</i>)	Late SD	20
Vernalization detection		
<i>VRN1</i> (<i>VERNALIZATION INSENSITIVE 1</i>)	Late LD, SD after vernalization	21
<i>VRN2</i> (<i>VERNALIZATION INSENSITIVE 2</i>)	Late LD, SD after vernalization	21
Hormonal regulation		
Cytokinins		
<i>AMP1</i> (<i>ALTERED MERISTEM PROGRAM 1</i>)	Early SD	22
Ethylene		
<i>EIN</i> (<i>ETHYLENE INSENSITIVE</i>)	Late LD	23
<i>ETR1</i> (<i>ETHYLENE RESPONSIVE 1</i>)	Late LD	23
Gibberellins		
<i>GA1</i> (<i>GIBBERELLIN REQUIRING 1</i>)	Late SD	24
<i>GAI</i> (<i>GIBBERELLIC ACID INSENSITIVE</i>)	Late SD	24
<i>SPY</i> (<i>SPINDLY</i>)	Early LD	25
Brassinosteroids		
<i>DET2</i> (<i>DEETIOLATED 2</i>)	Late LD	20
Carbohydrate metabolism		
<i>ADG1</i> (<i>ADP GLUCOSE PYROPHOSPHORYLASE 1</i>)	Late SD	26
<i>CAM1</i> (<i>CARBOHYDRATE ACCUMULATION MUTANT 1</i>)	Late LD	27
<i>PGM</i> (<i>PHOSPHOGLUCOMUTASE</i>)	Late SD	28
<i>SEX1</i> (<i>STARCH EXCESS 1</i>)	Late SD	29
<i>SUN2</i> (<i>SUCROSE UNCOUPLED 2</i>)	Early LD	30
Floral repressors		
<i>EMF1</i> (<i>EMBRYONIC FLOWER 1</i>)	Early LD, SD	31
<i>EMF2</i> (<i>EMBRYONIC FLOWER 2</i>)	Early LD, SD	31
<i>CLF</i> (<i>CURLY LEAF</i>)	Early LD, SD	32
<i>TFL1</i> (<i>TERMINAL FLOWER 1</i>)	Early LD, SD	33
Other genes that regulate flowering time		
<i>EBS</i> (<i>EARLY BOLTING IN SHORT DAYS =SPEEDY</i>)	Early LD, SD	34
<i>EFS</i> (<i>EARLY FLOWERING IN SHORT DAYS</i>)	Early LD, SD	35
<i>ELF1</i> (<i>EARLY FLOWERING 1</i>)	Early LD, SD	36
<i>ELF2</i> (<i>EARLY FLOWERING 2</i>)	Early LD, SD	36

<i>ELG (ELONGATED)</i>	Early LD, SD	37
<i>ESD1 (EARLY IN SHORT DAYS 1)</i>	Early LD, SD	38
<i>ESD4 (EARLY IN SHORT DAYS 4)</i>	Early LD, SD	38
Ectopic expression of wild type gene causes late flowering		
<i>CCA1</i>	Late LD, SD	39
<i>LHY (LATE ELONGATED HYPOCOTYL)</i>	Late LD	40
Ectopic expression of wild type gene causes early flowering		
<i>AG (AGAMOUS)</i>	Early LD, SD	41
<i>API (APETALA 1)</i>	Early LD, SD	42
<i>FPP1 (FLOWERING PROMOTING FACTOR 1)</i>	Early LD	43
<i>LFY (LEAFY)</i>	Early LD, SD	44

Table 1.1 Genes involved in the control of flowering time in *Arabidopsis thaliana*..

Abbreviations: LD, long days; SD, short days.

References: 1) Putterill *et al.*, 1995 2) Martinez-Zapater *et al.*, 1994 3) Lin *et al.*, 1996 4) Grbic and Bleecker, 1996 5) Macknight *et al.*, 1997 6) Koornneef *et al.*, 1994 7) Sanda and Amasino, 1996b 8) Clark and Dean, 1994 9) Araki *et al.*, 1997 10) Lee *et al.*, 1994b 11) Ray *et al.*, 1996 12) Deng *et al.*, 1992 13) Pepper *et al.*, 1994 14) Zagotta *et al.*, 1996 15) Parks and Quail, 1991 16) Goto *et al.*, 1991 17) Ahmad and Cashmore, 1993 18) Whitelam and Harberd, 1994 19) Ahmad and Cashmore, 1996 20) Pepper and Chory, 1997 21) Chandler *et al.*, 1996 22) Chaudhury *et al.*, 1993 23) Ecker, 1995 24) Wilson *et al.*, 1992 25) Jacobsen and Olszewski, 1993 26) Lin *et al.*, 1988 27) Eimert *et al.*, 1995 28) Caspar *et al.*, 1985 29) Caspar *et al.*, 1991 30) Dijkwel *et al.*, 1997 31) Sung *et al.*, 1992 32) Goodrich *et al.*, 1997 33) Shannon and Meeks-Wagner, 1991 34) Hicks *et al.*, 1996b 35) Soppe *et al.*, 1997 36) Zagotta *et al.*, 1992 37) Halliday *et al.*, 1996 38) Simon and Coupland, 1996 39) Wang *et al.*, 1997 a and b 40) Schaffer *et al.*, 1998 41) Mizukami and Ma, 1997 42) Mandel and Yanofski, 1995 43) Kania *et al.*, 1997 44) Weigel and Nilsson, 1995

1.2.1.2.2.2 Early-flowering mutants

The first descriptions of early flowering mutants appeared after those describing late flowering mutants (Sung *et al.*, 1992; Zagotta *et al.*, 1992; Coupland *et al.*, 1993; Yang *et al.*, 1995). Most work so far has focused on early ecotypes, in which earlier flowering mutant phenotypes are less pronounced than late flowering ones, unless grown under certain environmental conditions such as short days.

Many early flowering mutants have additional pleiotropic effects on development, or are compromised in other processes such as light perception and transduction or hormone regulation. Over twenty early flowering mutants have been identified, and these are summarized in Table 1.1. The most dramatic of these, *emf1* and *emf2* (*embryonic flower*), appear not to have a vegetative phase of development, but produce only a few cauline leaves before flowering (Sung *et al.*, 1992). Several mutants involved in light perception are also early flowering. These include *hyl*, 2, 3 and 6 (*long hypocotyl*, Goto *et al.*, 1991; Reed *et al.*, 1993; Koornneef *et al.*, 1980, 1995; Halliday *et al.*, 1994; Whitelam and Harberd, 1994) and *pef1*, 2 and 3 (*phytochrome-signalling early flowering*, Ahmad and Cashmore, 1996). Mutants which affect the circadian rhythm can also be early flowering. The *elf3* (*early flowering*) mutant is arrhythmic in constant light (Hicks *et al.*, 1996a) while the early flowering *det1* and *copl* mutants have a shorter rhythm in the dark (Millar *et al.*, 1995). Mutants which disrupt hormone signal transduction pathways, and are early flowering, have been identified. The *ampl* (*altered meristem program*) mutant has elevated levels of cytokinins (Chaudhury *et al.*, 1993), while the *spy* (*spindly*) mutant has a phenotype similar to that of wild type when gibberellic acids have been exogenously applied (Jacobsen and Olszewski, 1993). Mutations in genes which regulate the expression of floral meristem identity genes can also lead to early flowering phenotypes. The *tfl1* (*terminal flower*) mutant flowers early, and causes determinate growth, where inflorescences are replaced by flowers, and the shoot terminates prematurely in the

formation of a terminal flower (Shannon and Meeks-Wagner, 1991). The *clf* mutant (*curly leaf*) has a number of phenotypes in addition to early flowering, including leaf curling and the homeotic transformation of floral organs (Goodrich *et al.*, 1997).

In addition to the mutants which have been characterized according to their other mutant phenotypes, and which happen also to be early flowering, there are a number of early flowering mutants where the function of the genes affected has not yet been determined. These include *elf1* and 2 (Zagotta *et al.*, 1992) and the daylength-insensitive mutants *esd1*, *esd4* (*early in short days*, Simon and Coupland, 1996) and *ebs* (*early bolting in short days*) (reviewed in Hicks *et al.*, 1996b).

1.2.1.2.3 Models for the genetic control of flowering time

It has recently been proposed that the transition to flowering is the developmental default state in the *Arabidopsis* shoot apical meristem (Yang *et al.*, 1995). There are two main pieces of evidence to support this hypothesis. First, *Arabidopsis* can flower in the dark if sucrose is provided, with no environmental stimuli, and nearly all late flowering mutants show wild type flowering time in these conditions (Rédei *et al.*, 1974; Araki and Komeda, 1993a; Madueño *et al.*, 1996). Second, mutants which show almost no vegetative development have been isolated. Mutations in *EMF1* and *EMF2* (*EMBRYONIC FLOWER*) result in plants with cotyledons but no rosette leaves and an inflorescence that in weak mutant alleles contains a small number of sessile cauline leaves and a floral structure. Strong alleles produce no leaves and terminate in a carpelloid structure (Sung *et al.*, 1992). In the *emf* mutants expression of the meristem identity gene *AP1* (*APETALA1*), which is normally confined to the developing inflorescence, was detected in vegetative tissues, including the cotyledons, hypocotyl and, in weak mutants, the sessile leaves (Chen *et al.*, 1997). The *EMF* genes are thought to encode a floral repressor, or a promoter of vegetative development, which suppresses the default flowering pathway and allows vegetative

development to occur. The activity of this repressor would then be predicted to decrease throughout plant development allowing the transition to reproductive growth (Chen *et al.*, 1997). The role of the flowering time genes would be to promote or repress this central flowering suppressor (Martínez-Zapater *et al.*, 1994; Weigel and Nilsson, 1995). The *emf1* and *emf2* mutations are epistatic to *co* and *gi*, which delay flowering, suggesting that the *CO* and *GI* genes act to repress EMF function, allowing flowering to occur in response to inductive long days, and that in their absence the delay in flowering of the *co* and *gi* mutants is caused by increased EMF activity (Yang *et al.*, 1995; Castle and Sung, 1995). The timing of the transition to flowering is controlled by the flowering time genes, at least some of which promote the expression of the floral meristem identity genes (Simon *et al.*, 1996). As previously discussed, a number of these flowering time genes have been identified by mutation. The first to be characterized were mutants which delay flowering (Koornneef *et al.*, 1991). No mutants which completely prevent the transition to flowering have been recovered in *Arabidopsis*, suggesting that there may be redundancy between flowering time genes. The phenotype of these late flowering mutants, and the phenotype of plants carrying two mutations, allowed the genes involved in promoting flowering to be placed in at least three different genetic pathways (Koornneef *et al.*, 1991, 1998). The combination of two mutations which affect the same genetic pathway did not lead to an enhancement of phenotype: the double mutants flowered as late as the later of the two single mutants. The combination of two mutations which affect different genetic pathways, on the other hand, led to an enhancement of phenotype, with the double mutant flowering much later than either single parent as a consequence of the inactivation of more than one partially redundant pathway (Koornneef *et al.*, 1991; Coupland, 1995).

The first of these pathways is the environmental response pathway, also sometimes referred to as the long day promotion pathway. Mutations in genes involved in this pathway cause

the plant to flower late, but with reduced sensitivity to daylength and vernalization. The second is the autonomous pathway. Mutations in genes involved in this pathway cause plants to flower late but the mutants retain a strong response to daylength and vernalization (Koornneef and Peeters, 1997). The third pathway involves the promotion of flowering through gibberellins. GA mutants flower late primarily in SDs (Wilson *et al.*, 1992). As more flowering time mutants are analyzed particularly early flowering mutants, which may be compromised in floral repression, and vernalization-insensitive mutants, the model becomes more complex and the interactions between the mutations less well understood. However, these three partially redundant promotion pathways, identified by the phenotypic analysis of the twelve "classical" late flowering mutants and those mutations which affect GA responses, still form the basis of the model to explain the transition to flowering (Koornneef *et al.*, 1991, 1998; Figure 1.1).

1.2.1.2.3.1 The environmental response pathway

The transition to flowering in *Arabidopsis* is promoted by long days and vernalization. Mutations in the genes which are proposed to act in this pathway cause daylength- and vernalization-insensitive late flowering, although they do respond to very long cold treatments. This pathway is also called the long-day promotion pathway, because mutations in this group delay flowering under LDs but not SDs, suggesting that the pathway may promote flowering specifically in response to LDs. The *co* mutant was found to be in this group, in addition to *fd*, *fe*, *fha*, *ft*, *fwa* and *gi* (Koornneef *et al.*, 1991). However, although these mutants have been grouped together they still display some differences in particular environmental conditions. Under SD conditions, *co*, *gi* and *fha* mutants flower at the same time as wild type plants, whereas *ft*, *fwa*, *fd*, and *fe* maintain some photoperiodic response, flowering slightly later than wild type plants (Koornneef *et al.*, 1991). Furthermore, providing sucrose to the shoot apex in complete darkness was able to rescue the late

flowering phenotype of all the late flowering mutants except for *ft* and *fwa* suggesting that the genes in this group may be involved in more than one developmental response (Madueño *et al.*, 1996). Double mutant analysis has also revealed redundancy between *FT* and *FWA* and the floral meristem identity gene *LEAFY* (Ruiz-Garcia *et al.*, 1997), and possible overlapping functions for *FT* and *FE* and a gene involved in the autonomous flowering pathway, *FPA* (Koornneef *et al.*, 1998). Figure 1.1 shows the pathways involved in the promotion of flowering and some of the interactions between the genes in these pathways.

CO, the subject of this thesis, was the first of the group to be cloned. It was found to encode a protein with similarity to the GATA-1 family of transcription factors (Putterill *et al.*, 1995). The isolation of the gene and its initial characterization is described later in section 1.3.

Three other genes involved in long day promotion of flowering have recently been cloned. These are *FHA* and *FT*, which were originally characterized as mutations by Koornneef *et al.* (1991), and *LHY* (*LATE ELONGATED HYPOCOTYL*), which was recently isolated from a transposon-tagged population (Coupland, 1997; Schaffer *et al.*, 1998).

As described earlier, *FHA* was found to encode a blue light receptor, CRY2 (Guo *et al.*, 1998), suggesting that blue light acts through the long day pathway. The *FT* gene encodes a protein with homology to the meristem identity gene, *TFL1* (*TERMINAL FLOWER*) (Araki *et al.*, 1997). The *lhy* mutation, also mentioned earlier, was caused by the overexpression of a MYB transcription factor. As well as later flowering in LDs, this caused disrupted circadian clock function (Schaffer *et al.*, 1998).

1.2.1.2.3.2 The autonomous pathway

A pathway which promotes flowering independently of environmental stimuli was defined by another group of late flowering mutants which flower late, but are still responsive to daylength and vernalization. They flower later than wild type in both LDs and SDs, so unlike the mutants in the environmental pathway, they do not confer a day-neutral or close to day-neutral phenotype. The late flowering of these mutants can also be corrected by a period of vernalization. Mutants in this group, which affect the autonomous (or constitutive) pathway, are *fca*, *fpa*, *ld*, *fve* and *fy* (Koornneef *et al.*, 1991; Figure 1.1). Another late flowering mutant in this group with no obvious pleiotropic effects is *fld* (Sanda and Amasino, 1996a). This mutant was isolated in the Columbia background. Like the induced *ld* mutation, and the naturally occurring *FRI* allele, which is also proposed to act in the autonomous pathway, *fld* requires a dominant *FLC* allele in order to exhibit late flowering (Sanda and Amasino, 1996b).

Two genes, *ESD4* and *EFS*, which repress flowering have been proposed to act in the autonomous pathway. Mutations in these genes cause early flowering, and have been shown to be epistatic to the *fve* mutation that causes late flowering (Simon and Coupland, 1996; Soppe *et al.*, 1997).

Two genes in the autonomous pathway have been cloned: *LD* (*LUMINIDEPENDENS*) and *FCA*. *LD*, which was the first flowering time gene to be cloned, was isolated by T-DNA tagging (Lee *et al.*, 1994b). It encodes a putative transcription factor, with two bipartite nuclear localization domains, a glutamine-rich domain, which could act as a transcriptional activator, and a possible homeodomain (Aukerman and Amasino, 1996). The *FCA* gene encodes a protein with two RNA-binding domains and a WW protein interaction domain suggesting that it is involved in post-transcriptional regulation of transcripts. The gene itself is alternatively spliced, giving rise to four different transcripts (Macknight *et al.*, 1997).

1.2.1.2.3.3 The gibberellic acid promotion pathway

The transition to flowering in *Arabidopsis* is promoted by gibberellins, since mutations affecting either gibberellic acid biosynthesis (*gal*) or signal transduction (*gai*) are very late flowering in SDs (Koornneef and van der Veen, 1980; Wilson *et al.*, 1992). In addition, mutations in the *SPINDLY* (*SPY*) gene, which cause activation of GA-mediated responses in the absence of GA, cause early flowering (Jacobsen and Olszewski, 1993). This suggests that in SDs, the GA hormone is required for flowering. GA may compensate for the lack of the long day promotion pathway in SD non-inductive conditions. This was suggested by the observation that double mutants between *co*, defective in the LD pathway, and *gal* or *gai* flowered later than either parental line, and in LDs resembled SD-grown severe *gal* mutants (Putterill *et al.*, 1995; Figure 1.1).

The proposal that GA acts in a pathway independently of both the autonomous and environmental response pathways came from the observation that exogenous application of GA to late flowering mutants affecting either pathway was able to promote early flowering (Chandler and Dean, 1994). However, a possible interaction between the autonomous pathway and the GA promotion pathway was suggested from the results of a study of the *fve* mutant, which appeared to show symptoms of reduced GA levels or action (Martínez-Zapater *et al.*, 1995). In addition, a role for GA in the vernalization response has been proposed, from the finding that the severe *gal-3* mutant in SDs cannot respond to vernalization (Chandler *et al.*, 1996). Although the interaction between the GAs, the constitutive promotion of flowering and the vernalization response remains unclear, a candidate target gene, *FPP1* (*FLOWERING PROMOTING FACTOR*), that was proposed to be part of the GA promotion pathway was recently isolated (Kania *et al.*, 1997). Overexpression of this gene in transgenic plants caused early flowering and increased sensitivity to GAs.

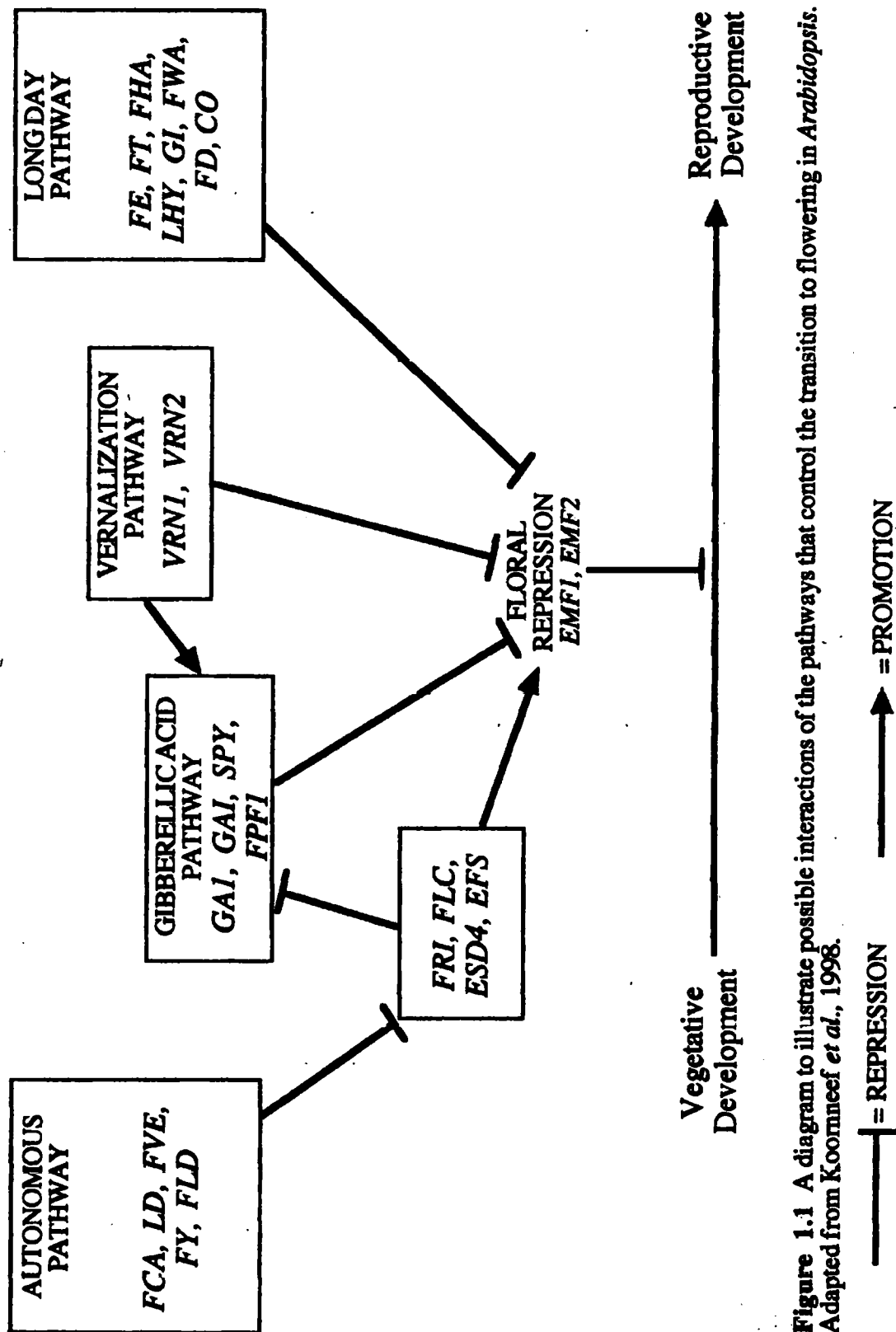


Figure 1.1 A diagram to illustrate possible interactions of the pathways that control the transition to flowering in *Arabidopsis*. Adapted from Koomneef *et al.*, 1998.

1.2.1.3 The role of *CO* in the promotion of flowering

The role of the *CO* gene in the long day promotion pathway and its relationship with other genes in the same pathway is starting to emerge. The following approaches have been used:

- the construction of transgenic plants ectopically expressing the *CO* gene,
- introduction of the *CO* transgene into other flowering time or meristem identity mutants,
- the analysis of expression of putative target genes of *CO* after induction of *CO* expression,
- the analysis of levels of *CO* expression in another mutant affected in the LD promotion pathway.

The levels of the *CO* mRNA in all tissues and developmental stages of wild type plants are extremely low, but approximately three times higher in LD-grown plants than those grown in SDs (Putterill *et al.*, 1995; Guo *et al.*, 1998; Suárez-López and Coupland, 1998). The very low level of expression under SDs is consistent with the phenotype of the *co* mutant, which is late flowering in LDs, but flowers at approximately the same time as wild type in SD. The expression analysis led to the suggestion that *CO* is upregulated in response to LDs, and that transcriptional regulation is a component of the LD pathway. Furthermore, analysis of transgenic plants suggested that increasing the expression of *CO* under SDs might be sufficient to promote early flowering under these conditions. Transgenic *co-2* mutant plants carrying multiple copies of a T-DNA containing *CO* flowered earlier than wild type in both LDs and SDs (Putterill *et al.*, 1995). To test the effect of *CO* mRNA levels on flowering time transgenic plants were made that contained a *CO* gene expressed from the cauliflower mosaic virus (CaMV) 35S promoter. These transgenic plants flowered very early in both LDs and SDs (Simon *et al.*, 1996). However, the 35S:*CO* plants still showed slight daylength sensitivity, flowering slightly later in SDs than in LDs, and displayed differences in the node at which bolting occurs (Igeño *et al.*, unpublished results).

In both LDs and SDs the transgenic plants made rosette leaves, so are not as early flowering as the most extreme early mutants, such as *emf1* and 2. Very early flowering must require additional factors. Recently, it was observed that introducing the *tfl1* mutation into the 35S:*CO* background caused earlier flowering than either parental line, suggesting that in the 35S:*CO* transgenics, *TFL1* is delaying the transition to flowering (Igeño *et al.*, unpublished results).

The 35S:*CO* transgene was also introduced into an *fca-1* mutant background to test whether increasing the activity of the long day promotion pathway might compensate for the loss of the autonomous pathway. The 35S:*CO fca-1* plants flowered slightly later than 35S:*CO*, but much earlier than the single *fca-1* mutant, indicating that while overexpressing *CO* can almost completely compensate for the loss of the autonomous pathway, *FCA* is still required for the very early flowering of 35S:*CO* plants. This suggests that at some point the pathways interact to promote flowering (Igeno *et al.*, unpublished results).

Inducible *CO* expression was also achieved by constructing a translational fusion between *CO* and the rat glucocorticoid receptor, under the control of the CaMV 35S promoter (Simon *et al.*, 1996). This construct (35S:*CO:GR*) was introduced into the late flowering *co-2* mutant. The resulting protein was inactive in the absence of a steroid ligand, but addition of the hormone dexamethasone restored the activity of the protein and resulted in rapid flowering in both LDs and SDs, suggesting that *CO* can promote flowering at any time in development (Simon *et al.*, 1996). This system was used to analyze the expression of floral meristem identity genes in response to *CO* activation.

Mutations in floral meristem identity genes disrupt floral development, leading to the production of shoots where flowers would normally form in wild type plants. As mentioned previously, the most well characterized of these genes are *LEAFY* (*LFY*), *APETALA1*, *APETALA2* (*AP1*, *AP2*) and *CAULIFLOWER* (*CAL*) (Haughn *et al.*, 1995). In addition,

mutants with the opposite effect have been characterized; the *tfl1* and *tfl2* mutants convert shoots to single flowers and terminate the main stem prematurely with a terminal flower (Shannon and Meeks-Wagner, 1991; Hicks *et al.*, 1996b). The relationship between mutants affecting flowering time and those affecting floral meristem identity has been studied genetically by making double mutants and by studying expression levels of these genes in different mutant backgrounds. When *co-2* was combined with a weak *lfy* allele, for example, the phenotype was strongly enhanced (Putterill *et al.*, 1995) suggesting a close relationship between the two genes.

To study the effect of *CO* expression on a number of floral meristem identity genes the 35S:*CO*:*GR* transgenic plants were treated with dexamethasone (dex) to activate *CO* function, and the expression of *LFY*, *API* and *TFL1* studied by *in situ* hybridization in response to this treatment (Simon *et al.*, 1996). *LFY* and *TFL1* transcription was induced within 24 h after application of dex, regardless of whether the plants were grown in LDs or SDs. *LFY* expression was also induced within this time in plants shifted from SDs to LDs, indicating that *LFY* can respond as quickly to induced *CO* expression as it can to inductive LDs. *API* transcript, on the other hand, was not detected until 72 h after treatment with dexamethasone in LDs, or after shifting plants from SDs to LDs, which is consistent with the genetic data which suggests that *API* acts after *LFY* (Weigel and Nilsson, 1995; Mandel and Yanofsky, 1995). Furthermore, after induction of *CO* activity with dex in SDs, the *API* transcript was not detectable until 120 h after treatment. This indicated that *API* is transcribed quicker in response to LDs than in response to *CO*. These data suggest that *CO* acts in a pathway that activates *LFY* and *TFL1* transcription in response to LDs, and that an additional LD pathway is required for the activation of *API*. Whether *CO* acts directly upon the *LFY* and *TFL1* promoters to activate transcription is as yet unknown.

The *CO* transcript was less abundant in the late flowering, daylength-insensitive *fha* mutant (Guo *et al.*, 1998). *FHA* was shown to encode the blue light receptor cryptochrome 2 suggesting a link between regulation of the long-day promotion pathway by blue light and *CO* expression. *CO* transcription was also upregulated by ectopic expression of *FHA*, which also caused early flowering in SDs. This suggested that *FHA* acts before *CO* in the long day promotion pathway and is required for the increased expression of *CO* and earlier flowering in LDs compared to SDs (Guo *et al.*, 1998, Suárez-López and Coupland, 1998).

1.3 Isolation of the *CONSTANS* gene

The *CO* gene had previously been genetically mapped to position 16.4 cM on the top arm of chromosome five, approximately 1.6 cM from the *CHALCONE SYNTHASE* gene (*CHS*) (Koorneef *et al.*, 1983). In 1990 a chromosome walk was initiated by Putterill *et al.* from *CHS* and other RFLP markers which, from the RFLP maps available were predicted to be close to the *CO* gene (Chang *et al.*, 1988; Nam *et al.*, 1989). A YAC contig extending over approximately 1700 kb and containing 67 YAC clones was built during the isolation of the gene (Putterill *et al.*, 1993). *CO* was located first to a 300 kb region within the YAC contig and subsequently this was narrowed down to a physical distance of approximately 48 kb, covered by three overlapping cosmid clones (Putterill *et al.*, 1995). This was possible because of the presence of the *alb2* mutation 1.6 cM from *CO*, which allowed the positive selection of plants in which recombination events had occurred on both sides of *CO*. Plants in which no recombination event had occurred between *alb2* and *co* were homozygous for the lethal *alb2* mutation. The three cosmid clones were introduced into the *co-2* mutant by *Agrobacterium*-mediated root transformation and two of them successfully complemented the mutant phenotype. This further narrowed down the region proposed to contain the *CO* gene to approximately 7 kb, defined as the region in which the two cosmids

overlapped. Using subclones from this region a number of cDNA libraries were screened to try to isolate a cDNA for the *CO* gene. As this was unsuccessful the 7 kb region was sequenced and searched using the GRAIL programme (Uberbacher and Mural, 1991) for putative ORFs (Open Reading Frames). The two longest putative ORFs were found on the same strand separated by only 51bp. The final proof that these two ORFs corresponded to exons of the *CO* gene came from sequence analysis of the ORFs, amplified in duplicate by PCR from three of the *co* alleles. All three contained sequence changes which would be predicted to alter the amino acid sequence of the protein. Two of these were published in Putterill *et al.*, 1995. The *co-1* mutation is a nine base pair deletion which precisely removes three amino acids, leaving the rest of the protein in frame. The *co-2* allele contains a point mutation predicted to cause an arginine to histidine substitution (discussed in detail in Chapter 3, Section 3.2.3).

1.4 The CONSTANS protein has a novel cysteine-rich domain which is predicted to be able to form two zinc fingers

When the *CONSTANS* gene was first sequenced in 1993, the databases available at the time were searched for proteins with homology to the predicted CONSTANS protein. A FASTA search comparing the CONSTANS protein to sequences in GenBank detected no significant homologies and a PROSITE search failed to detect any motifs within the protein (Putterill *et al.*, 1995). However, analysis of the protein sequence by eye revealed an arrangement of cysteine residues near the amino terminus of the protein $[(C-X_2-C-X_{16}-C-X_2-C)]_2$ (C=cysteine, X=any residue) characteristic of the zinc-binding motifs found in a number of transcription factors.

1.4.1 Zinc finger proteins

The most common function of zinc-binding motifs similar to the one present in CO is to bind DNA, and they are characterized by having a number of cysteine and/or histidine residues which interact with zinc ions to stabilize the structure of a small, autonomously folded protein domain (reviewed in Klug and Schwabe, 1995).

The first "zinc finger" containing protein to be described was the TFIIIA transcription factor isolated from the frog *Xenopus laevis* (Pelham and Brown, 1980). It was found to contain nine repeats of a 30 amino acid motif with conserved cysteine, histidine and hydrophobic residues (Φ -X-C-X_{2,5}-C-X₃- Φ -X₅- Φ -X₂-H-X_{2,5}-H, where X = any amino acid, Φ = a hydrophobic residue, C = cysteine and H = histidine, the classical C₂H₂ zinc finger; Miller *et al.*, 1985). Each motif was found to form a discrete DNA-binding domain with four zinc-binding residues in a tetrahedral arrangement around a central zinc ion. Since then this motif and a number of variants have been found to be widespread in eukaryotic DNA binding proteins. At least ten distinct zinc-binding domains have been studied in detail, and in addition, a number of novel putative zinc-binding motifs containing variations on the characteristic spacing of cysteine and/or histidine residues have been identified (reviewed in Klug and Schwabe, 1995 and in Schmiedeskamp and Klevit, 1994). Some of the well-characterized zinc-binding domains and a few novel motifs have been found in plant proteins (Table 1.2).

There are to date at least 36 plant proteins which have been found to contain the classical C₂H₂ TFIIIA-type zinc-finger (Meissner and Michael, 1997; Table 1.2). The majority of these contain either one or two tandem copies of the C₂H₂ motif, which is in contrast to the multiple C₂H₂ zinc fingers found in animal proteins (Jacobs, 1992). Eight members of a one-fingered family (AtZFP) were isolated from *Arabidopsis* (Tague and Goodman, 1995) and the *Arabidopsis SUPERMAN* gene, which is involved in negatively regulating B-function floral organ identity genes, was also found to encode a single-finger protein only

distantly related to the AtZFP family (Sakai *et al.*, 1995). The C₂H₂ two-fingered plant proteins include the floral organ-specific *Petunia* EPF-1 and EPF-2 families (Takasuji *et al.*, 1994) and the Wzf1 ZFP from wheat, which binds to a motif in the histone H3 promoter (Sakamoto *et al.*, 1993). From pea petals a C₂H₂ ZFP cDNA was isolated by PCR using degenerate primers designed to the coding sequence for the *Petunia Epfl* gene zinc fingers (Michael *et al.*, 1996) and this strategy was also used to isolate five two-fingered and the first example in plants of a three fingered ZFP cDNA from *Arabidopsis* (Meissner and Michael, 1997).

The human glucocorticoid receptor, the first member of a large family of nuclear hormone receptors, was cloned at around the same time as the identification of the zinc finger motif in TFIIIA (Hollenberg *et al.*, 1985). The DNA binding domain in this family is characterized by a conserved spacing of cysteines, reminiscent of the classical zinc finger (Kumar *et al.*, 1987). Unlike the modular C₂H₂ fingers, however, the two fingers of the nuclear hormone receptors are folded together to form a single structural domain. There is an example in plants of a protein with homology to a number of nuclear hormone receptors: The gene encoding the ES43 protein was isolated from barley by degenerate PCR, using primers designed to the conserved DNA binding domain of the animal receptors (Speulman and Salamini, 1995).

The C₂-H-C₅ LIM domain was first identified in three animal proteins: LIN-11 from *C. elegans* (Freydt *et al.*, 1990), ISL-1 from rat (Karlsson *et al.*, 1990) and MEC-3 from *C. elegans* (Way and Chalfie, 1988). Since then it has been found in over 90 animal proteins (Dawid *et al.*, 1995), but to date in only three yeast and one plant protein. Each LIM domain consists of two zinc fingers separated by two amino acids and each protein can have one, two or three tandem LIM domains. The pollen-specific PLIM-1 (formerly SF3) from Sunflower has two LIM domains which have recently been implicated in DNA binding (Baltz *et al.*, 1996) although other LIM domains have been proposed to have a role

in protein-protein interaction (Schmeichel and Beckerle, 1994). The spacing of zinc-binding residues in the LIM domain is similar to that found in the GATA-type zinc fingers (Sánchez-García and Rabbitts, 1994; Table 1.2) except that the zinc binding residues in the GATA-type fingers are all cysteines. In more recent database searches for genes with homology to *CONSTANS*, the LIM domain of the putative *C. elegans* transcription factor CeLIM-4, (Accession U72348; Hobart *et al.*, unpublished results) was identified because of the conserved spacing of some of the cysteines and a number of other residues (Chapter 3).

Two other cysteine-rich domains which possibly have a role in protein-protein interactions are the C₃-H-C₄ RING finger (**REALLY INTERESTING NEW GENE**; Freemont *et al.*, 1991) and the C₄-H-C₃ PHD finger (**PLANT HOMEODOMAIN**; Aasland *et al.*, 1995). The RING finger has been found in nearly 80 diverse proteins from many species (Saurin *et al.*, 1996) and to date four plant genes reportedly contain RING finger-like domains. The most well characterized of these is the *Arabidopsis* COP1 protein, involved in repression of the photomorphogenic pathway in the dark and in a number of light responses (Deng *et al.*, 1992). The two recently isolated *Arabidopsis* proteins each containing a RING finger are the ATL2 protein, putatively involved in the auxin early response, and the A-RZF protein which appears to have a role in seed development. The former was isolated from a cDNA expression library by screening for clones which were toxic to yeast (Martínez-García *et al.*, 1996) and the latter was identified from a screen for mutants defective in diacyl-glycerol acyltransferase activity (Zou and Taylor, 1997). The fourth example of a gene encoding a putative RING finger protein was found in a salt-induced cDNA in salt-tolerant alfalfa (Winicov, 1993). The PHD finger, as its name suggests, was identified in two plant proteins which also contain the well characterized DNA binding homeodomain. Many other examples of this zinc finger domain have been subsequently found in animal homeodomain proteins (Aasland *et al.*, 1995). The first plant proteins to define this domain

were the *Arabidopsis* HAT3.1 protein, putatively involved in root development (Schindler *et al.*, 1993) and the maize *Zmhox1a* protein, which regulates the expression of the *Shrunken* gene (Bellmann and Werr, 1992). Two homeodomain proteins involved in the transcriptional regulation of pathogen defence-related genes also contain a PHD finger: PRHP from parsley and PRHA from *Arabidopsis*. The role of the PHD finger in all these proteins is unclear but it could be involved in the stability or specificity of the protein-DNA interactions mediated by the DNA-binding homeodomain (Korfhage *et al.*, 1994).

One cysteine-rich motif which so far has only been found in plant transcription factors is the Dof domain (DNA-binding with one finger). It consists of a C-X₂-C-X₂₁-C-X₂-C motif predicted to form a single zinc finger and was first identified in the maize MNB1a DNA-binding protein (Yanagisawa, 1995). Like the LIM domain and the GATA motifs (discussed later) it appears to have a role both in DNA binding and in protein-protein interaction through the formation of homo- and heterodimers with other Dof proteins (Yanagisawa, 1997). The gene encoding MNB1a (now called Dof1) is part of a multigene family in maize, with at least 10 members, and cDNAs which appear to encode proteins containing a Dof domain have also been isolated from *Arabidopsis*, pumpkin and tobacco (Yanagisawa, 1996; Zhang *et al.*, 1995). One of the *Arabidopsis* cDNAs encodes a Dof domain-containing protein, OBP1, which enhances the binding of the bZIP transcription factor, OBF, to the octopine synthase promoter (Zhang *et al.*, 1995). Another Dof protein, NtBBF, was isolated from tobacco (DePaolis *et al.*, 1996). This protein binds to the promoter of the plant oncogene, *rolB* and regulates its expression in the meristem.

There are several other examples of novel cysteine-rich motifs which have been found only in plant proteins; the ABF1 and ABF2 proteins from wild oat, which bind to the α -*Amy2* promoter, defined a class of plant zinc finger proteins with a variation on the C₂H₂ motif (Rushton *et al.*, 1995). This motif, in which the spacing between the zinc-binding residues is different from the spacing in the classical TFIIIA-like C₂H₂ zinc fingers, was also found

in a number of homologues from *Arabidopsis*, *Brassica campestris* and rice and in the DNA-binding protein SPF1 from sweet potato (Ishiguro and Nakamura, 1994).

One of the first plant zinc finger proteins to be characterized was the 3AF1 DNA binding protein from tobacco, which interacts with a constitutive element within the light-responsive promoter of pea *rbcS-3A* (Lam *et al.*, 1990). The protein contains two homologous stretches of approximately 100 amino acids, with a unique HC₂HC arrangement of potential zinc-binding ligands.

The light responsive promoter of pea *rbcS-3A* has been studied extensively and a number of conserved sequences, in addition to the constitutive element bound by 3AF1, have been identified (Lam *et al.*, 1990). These sequences have in common the motif T/A GATA G/A, the so-called GATA motif, which is also present in a number of other plant light-responsive gene promoters (Arguello-Astorga and Herrera-Estrella, 1996). This conserved motif was first identified as the target site for the interaction of a family of DNA binding proteins with the promoters of genes involved in hematopoiesis in vertebrates (Evans and Felsenfeld, 1989). These DNA-binding proteins, known as GATA factors after their DNA target site consensus, have in common a cysteine rich domain, C-X₂-C-X₁₇-C-X₂-C, which can form a zinc finger (Harrison, 1991). Animal GATA factors often contain two zinc fingers of this type. GATA factors have been found in a wide range of vertebrates, where they regulate the expression of genes involved in the development of red blood cells, the heart and the gut (Simon, 1995). Genes encoding zinc finger proteins with homology to the vertebrate GATA factors have subsequently been found in invertebrates, filamentous fungi, yeast and plants (reviewed in Teakle and Gilmartin, 1998a). The fungal GATA factors, such as AreA from *Aspergillus nidulans* (Kudla *et al.*, 1990) and NTT2 from *Neurospora crassa* (Fu and Marzluf, 1990), which contain a single zinc finger, are typically involved in nitrogen-regulated gene expression. Recently, however, the genes encoding the GATA-like factors WC1 (WHITE COLLAR) and WC2 from *Neurospora crassa* (Ballario *et al.*, 1996; Linden

Protein	Sequence
TFIIIA-like	CX₂C X ₁₂ HX₃H CX₄C X ₁₂ HX₃H
Glucocorticoid receptor-like	CX₂C X ₁₃ CX₂C X ₁₅ CX₅C X ₉ CX₂C X ₄ C
RING finger	CX ₂ C X ₁₂ CXHX ₂ CX ₂ C X ₁₁ CX ₂ C
PHD finger	[CX₂C X ₁₃ CX₃C] X ₄ [HX₂C X ₁₈ CX₂C]XC
DOF	CX₂C X ₂₁ CX₂C
ABF1/2	CX₄₋₅C X ₂₂₋₂₃ HXH
3AF1	H X ₁₂ CX ₂ C X ₁₄ H X ₁₂ C
LIM	[CX₂C X _{17±1} HX₂C] X ₂ [CX₂C X _{17±1} CX₂C]
GATA-1	[CX₂C X ₁₇ CX₂C] X ₃₀ [CX₂C X ₁₇ CX₂C]
CONSTANS	[CX₂C X ₁₆ CX₂C] X ₁₉ [CX₂C X ₁₆ CX₂C]

Table 1.2 Potential metal-binding residues in zinc-finger-like proteins compared to CONSTANS. The brackets indicate individual zinc fingers. Pairs of potential metal binding ligands in the zinc-binding motifs related to the "classical" TFIIIA zinc finger are shown in bold. Adapted from Sanchez-Garcia and Rabbitts, 1994 and Klug and Schwabe, 1995. C = cysteine, H = histidine, X = any residue.

TFIIIA (Pelham and Brown, 1980).
 Glucocorticoid receptor (Kumar *et al.*, 1987).
 RING finger (Lovering *et al.*, 1993).
 PHD finger (Schindler *et al.*, 1993).
 DOF (Yanagisawa, 1995).
 ABF1/2 (Rushton *et al.*, 1995)
 3AF1 (Lam *et al.*, 1990)
 LIM (Sanchez-Garcia and Rabbitts, 1994).
 GATA-1 (Omichinski *et al.*, 1993)

and Macino, 1997) have been shown to be involved in the control of blue light signal transduction, which might suggest that in plants as well as in fungi, the GATA motifs found in some light regulated promoters could be the target sites for GATA-like zinc finger proteins. In plants the first gene encoding a GATA-like single zinc finger protein, Ntl1(nit-2-like1) was isolated from tobacco by cross-hybridization to the *N. crassa nit-2* gene. NTL1 is putatively involved in both nitrogen- and light-regulated gene expression (Daniel-Vedele and Caboche, 1993). Recently, four cDNAs, corresponding to 4 different genes, were isolated from *Arabidopsis* based on their homology to GATA-encoding genes from animals and fungi (Teakle and Gilmartin, 1998a and b). The proteins they encode, GATA-1, -2, -3 and -4, have a single zinc finger of the C-X₂-C-X₁₈-C-X₂-C type characteristic of both the tobacco NTL1 factor and a subset of fungal GATA factors, including WC1 and WC2, implicated in blue light signal transduction.

1.4.1.1 Comparison of the CO protein with other zinc finger proteins

The spacing of some of the cysteine residues in the CONSTANS protein (C-X₂-C-X₁₆-C-X₂-C) is reminiscent of the GATA-1 family of transcription factors (C-X₂-C-X₁₇-C-X₂-C, Trainor *et al.*, 1990; Figure 1.2). The CONSTANS protein contains two such domains separated by 19 amino acids. However, apart from the conserved spacing of cysteine pairs there is little other direct homology to the GATA-1 transcription factor from plants or animals although there is one feature which is reminiscent of GATA-1. When the two proposed zinc fingers from CO are aligned they display considerable homology to each other, with 46% identical and 86% identical or similar amino acids. The homology is particularly apparent at the carboxyl side of each finger (Figure 1.2) where there are a number of basic amino acids and both regions have a net positive charge. These regions in GATA-1 transcription factors are also basic, highly conserved and implicated in DNA binding (Romain *et al.*, 1993) and the formation of homodimers (Crossley *et al.*, 1995).

First zinc finger CDTCRSNACTVYCHADSAYLCMSCDAAQVHSANRVASRIHKVRVRV
 |:| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Second zinc finger CESCERAPAAFLCEADDASLCTACDSEVHSANPLARRHQVRPI

CO CDTCRSNACTVYCHADSA-YLCMSCD--AQVHSANRVASRHKRVRV-----CESC
| : | :: | :: : : : || | : | : : : : | | : | : : | : |
hgATA1 CVNCGATATPLWRRDRTHGYLNCACGLYHKMNGQNRPLIRPKRLIVSKRAGTQCTNC

CO ERAPAAFL-CEADDASLCTACDSEVHSANPLARRHQVPVI
: : : : : : | : : : : | : | :
hgATA1 QTTTTTLWRRNASGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNRKASGKGKKKR

B) Comparison of the zinc finger region of CO with that of hGATA1 (Romain *et al.*, 1993) using the FASTA program, as in A. Horizontal lines indicate gaps introduced to maximize alignment.

Adapted from Putterill *et al.* (1995).

Table 1.2 shows a comparison of the cysteine pairs in the CO protein with the putative metal-binding residues in the nine distinct classes of zinc-finger-like motifs which have been found in plant proteins and which have been described here. The spacing of the cysteine pairs within each proposed finger of the CO protein is similar to both the GATA-1 and the LIM domain spacing, but the lack of additional homology and the presence of extra cysteine and histidine residues close to the cysteine pairs (discussed in Chapter 3) suggests that CO is a member of a GATA-1-related but distinct family of putative zinc-binding proteins.

1.5 Background and summary

From May, 1990, the author was involved in the isolation of the *CONSTANS* gene under the supervision of Dr. Joanna Putterill and Dr. George Coupland. When the complementation of the *co-2* mutant was in progress the author initiated the analysis of the seven *co* mutants as an independent thesis project under the supervision of Dr. George Coupland.

This thesis describes the phenotypic and molecular analysis of the seven *co* mutants, the isolation of a new deletion allele and the identification of conserved domains of the protein thought to be important for CO function. The analysis of structural conservation among homologues was extended to include the functional conservation of an orthologue of *CO* from *Brassica napus*.

2 Characterization of the phenotypes caused by seven *constans* mutant alleles

2.1 Introduction

The *constans* mutant was first isolated in a screen for "supervital" mutants after X-irradiation (Rédei, 1962; *co-1*). Late-flowering mutants appeared supervital because, compared to wild type, they showed an increase in fresh and dry weight and produced more progeny as a consequence of forming more flowers and seed pods. These changes were shown to be due to the reduced sensitivity of the *co* mutant to day length; it was late flowering compared to its wild type progenitor under long day inductive conditions, and flowered at approximately the same time as wild type under short days (Rédei, 1962; Figure 2.1). An additional six alleles at the *co* locus (designated *co-2* to *co-7*) were subsequently isolated by screening for late-flowering mutants in a population of plants mutagenized with EMS (Koornneef *et al.*, 1991).

This chapter describes the phenotypic characterization of plants homozygous or heterozygous for each of the seven *co* alleles, confirming and extending the analysis first published by Koornneef *et al.* (1991).

2.1.1 Backcrossing *co-1* into Landsberg *erecta*

The *co-1* allele was generated by Rédei (1962) after X-irradiation of seeds provided by Professor Laibach. However, this batch of seeds proved to be heterogeneous, and both the Columbia wild type and, after further irradiation and selection of desirable traits, Landsberg *erecta* were isolated from it (Rédei, 1992). Hence, unlike alleles *co-2* to *co-7*, the *co-1* mutant contained the wild type *ERECTA* gene and although it was originally described as

LD



Wild type

co-2

SD



Wild type

co-2

Figure 2.1 Wild type *Landsberg erecta* and the *co-2* mutant grown in long day (LD) and short day (SD) conditions. The LD grown plants are 28 days old and the SD grown plants are 42 days old.

having been isolated in Columbia (Rédei, 1962; Koornneef *et al.*, 1991), in the region around *co* it showed Landsberg polymorphisms with the RFLP markers *CHS2* (*CHALCONE SYNTHASE2*), Feinbaum and Ausubel, 1988) and g5962 (Hauge and Goodman, 1992).

Mutants for *erecta* are dwarfed, sturdy and upright with a compact stature and are convenient for genetic crossing and growth under greenhouse conditions (Torii *et al.*, 1996). This is particularly useful when analyzing large, late flowering plants. *ERECTA* plants carrying the *co-1* allele grow extremely tall, have thin stems and the plants tend to fall over and become entangled with neighbouring plants.

The relationship between flowering time (FT) and leaf number (LN) differs in different genetic backgrounds (Hussein, 1968). *co-1* mutants in an *ERECTA* background had a higher LN at the same FT than *erecta* plants carrying the other alleles (Koornneef *et al.*, 1991; Table 2.1). The inflorescence stalk (bolt) of *co-1* mutants also elongated more rapidly than the other mutants making it appear artificially early flowering (Figure 2.2a).

The *co-1* mutant was crossed to Landsberg *erecta* in order to analyze the semidominance of the mutation (see section 2.3), and an F₂ population derived from this cross was also used to identify a *co-1* plant homozygous for *erecta*. To confirm that this individual carried the *co-1* mutation and was not derived from a seed contaminant, two pairs of PCR primers were designed which could be used to distinguish between *co-1* mutant and *CO* wild type DNA (primers co59 and co60, each used in combination with co49; Figure 2.2b; Chapter 7, Materials and Methods). This was made possible by the presence of a nine base pair deletion in the *co-1* allele (Putterill *et al.*, 1995). These primers were used in a PCR experiment using DNA extracted from the proposed *co-1/erecta* double mutant plus a number of control wild type *Ler* and *co-1* plants (Figure 2.2b). These experiments confirmed that the *co-1/erecta* plant was homozygous for both mutations.



Figure 2.2a The *co-1* mutant in Landsberg with the wild type *ERECTA* gene (left) compared to *co-1* in Landsberg *erecta* (right). The plants were grown in LDs and are 42 days old.

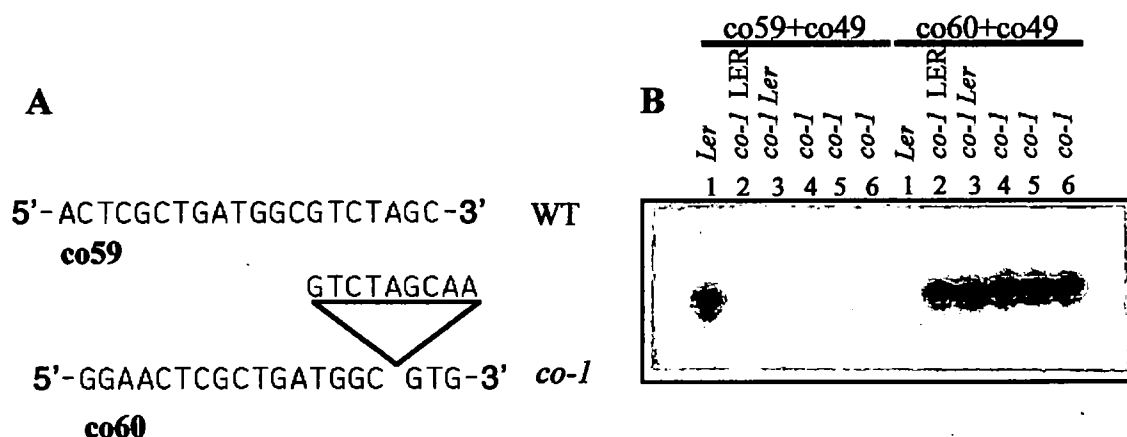


Figure 2.2b Panel A shows the primers designed to distinguish between wild type *CO* and the *co-1* mutant allele. Primer co59 was designed to anneal only to wild type DNA (or DNA from other mutant alleles) and co60 was designed to anneal only to *co-1* DNA. The 9 bp deletion in *co-1* is shown in blue.

Panel B shows a Southern blot of PCR products of reactions using primers co59 or co60 with primer co49, which would produce a 300 bp product in both cases. The blot was probed with the *CO* cDNA. Primer co59 only gave a product with wild type DNA, as expected, and primer co60 only gave a product with DNA from a number of *co-1* mutants, as expected. The PCR products from DNA extracted from the plants shown in Figure 2.2 are indicated (samples 2 and 3). The candidate *erecta/co-1* mutant (sample 3) was shown to contain the *co-1* mutation.

Once the genotype of this individual was confirmed, it was self fertilized, the seeds collected and the FT/LN determined alongside the other *co* mutants (see section 2.2.1). The *co-1* allele was backcrossed once to *Ler* and the individual chosen for phenotypic analysis of flowering time had a similar FT/LN ratio to the other *co* mutants and had the compact growth characteristic of *Landsberg erecta* (Figure 2.2a).

2.2 The phenotypes caused by seven *co* mutant alleles

The adult life cycle of *Arabidopsis* includes a vegetative phase, when leaves are produced in a spiral phyllotaxy on the flanks of the shoot apical meristem and give rise to a rosette, and a flowering phase when no new leaves are formed but flowers are produced by the primary meristem. During the flowering phase internode elongation also occurs to produce the tall inflorescence stem. Flowering time (FT) is measured either by counting the number of days from sowing to the first appearance of buds in the centre of the rosette or by counting the number of leaves (LN) formed during the vegetative phase (Koornneef *et al.*, 1991). The correlation between flowering time and leaf number has been measured for flowering time mutants and for wild type plants (Koornneef *et al.*, 1991). For many flowering time mutants, including *co*, the rate of leaf production is the same as for wild type, and the total number of leaves is often the preferred method for measuring differences in flowering time. Measuring the number of days to flowering can be complicated by non-genetic factors such as germination time. In this analysis both methods were used to measure flowering time.

2.2.1 Flowering time in long days

To analyze the changes in flowering time caused by the mutant alleles, twenty seeds of each mutant and of wild type *Landsberg erecta* controls were first stratified on damp filter paper at 4°C for five days to break dormancy and synchronize germination. The seeds were

then planted in a controlled environment room (CER) with a long-day photoperiod (16 h light, 8 h dark; LDs). All mutant lines had been backcrossed to Landsberg *erecta* at least twice, in order to reduce the likelihood that any other mutations caused by the initial mutagenesis would still be present. No visible phenotypes apart from late flowering were apparent in any of the mutants before or after backcrossing.

As described previously flowering time was measured as the time in days from sowing to the appearance of the buds in the centre of the rosette. The total number of leaves, divided into rosette and cauline leaves, was also counted after the plants had bolted. The results are summarised in Table 2.1, Figure 2.3 and 2.4. The weakest mutant allele is *co-4*, which caused mutants to flower after an average of 24.3 days with 11.3 leaves. The strongest is *co-3*, which caused mutants to flower after 39.9 days with 29.1 leaves (wild type, 18.9 days with 7.5 leaves). These two alleles, at each end of the allelic series, are shown with a wild type plant in Figure 2.5. The severity of the remaining alleles fall between *co-4* and *co-3* and all differ quantitatively with respect to flowering time and total leaf number.

The results obtained are broadly in agreement with those of Koornneef *et al.* (1991), and with one exception the order of severity of the alleles was the same. In this experiment, the *co-2* allele was found on average to be slightly earlier both in flowering time and total leaf number than the *co-6* allele, but their order was reversed in the data of Koornneef *et al.* (1991). There was also a quantitative difference between the results of Koornneef *et al.* and the results presented here, probably due to differences in growth conditions. In this experiment the mutants flowered slightly earlier and with fewer leaves than in the experiments of Koornneef *et al.*

2.2.2 Flowering time in short days

Rédei's original description of the *co-1* mutant indicated that it flowered slightly earlier and with fewer leaves than its wild type progenitor in short days (SDs) (Rédei, 1962).

Mutant	Flowering time		Leaf number		
	Days to flowering	Days to first open flower	Rosette	Cauline	Total
<i>Ler</i>	18.9±0.9	27.3±0.8	4.9±0.2	2.6±0.2	7.5±0.2
<i>co-4</i>	24.3±0.5	33.5±0.4	7.6±0.2	3.8±0.1	11.3±0.4
<i>co-5</i>	28.2±0.5	36.4±0.6	9.2±0.2	5.2±0.2	14.4±0.4
<i>co-2</i>	33.1±0.9	43.6±0.9	13.5±0.7	5.9±0.3	19.4±0.9
<i>co-6</i>	34.2±0.5	48.7±0.7	13.4±0.3	10.3±0.3	23.7±0.5
<i>co-1</i> Landsberg <i>erecta</i>	36.0±0.6	50.0±0.4	16.1±0.4	9.4±0.3	25.6±0.6
<i>co-7</i>	37.8±0.7	47.5±0.6	18.5±0.5	9.9±0.4	28.5±0.8
<i>co-1</i> Landsberg	38.6±0.8	50.3±0.5	17.9±0.3	12.3±0.2	30.2±0.4
<i>co-3</i>	39.9±0.6	53.4±0.7	17.7±0.8	11.4±0.5	29.1±1.2

Table 2.1 Flowering times and leaf numbers of the seven *co* mutants grown in LDs compared to wild type Landsberg *erecta* . The mutants are listed in order of increasing number of days to flowering.



Figure 2.4 The phenotypes of the seven *co* mutants, each carrying a different *co* mutant allele. All plants were grown in LDs and are 42 days old. Wild type Landsberg *erecta* was not included as it had senesced by this stage. The *co-1* mutant (top left) is in Landsberg *erecta* and the remaining six are in Landsberg *erecta*.

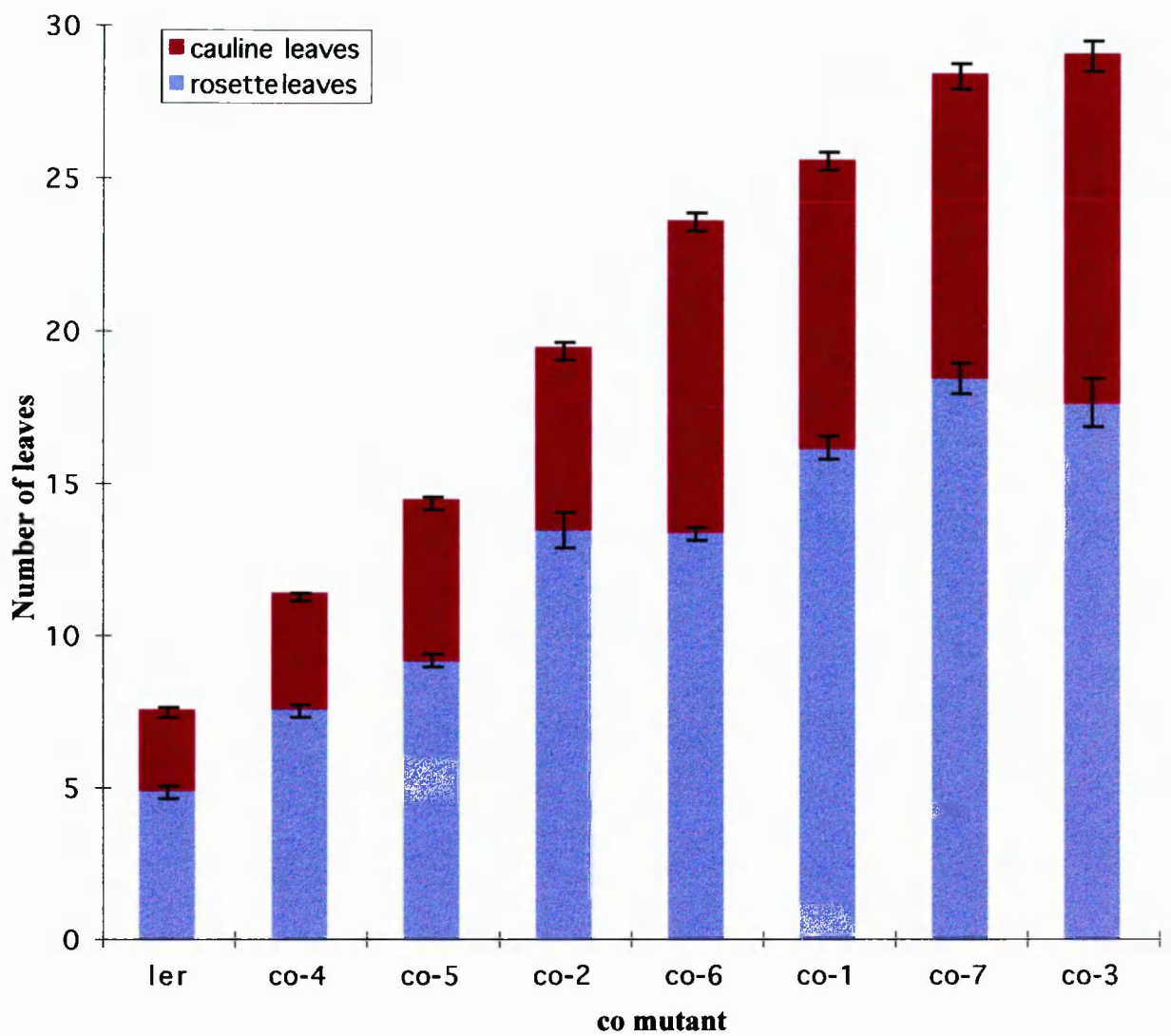


Figure 2.3 The flowering times, represented as the mean rosette and cauline leaf numbers of the seven *co* mutants compared to wild type Landsberg *erecta* in LDs. The mutants are shown in order of severity of phenotype. Standard error bars are shown in black.

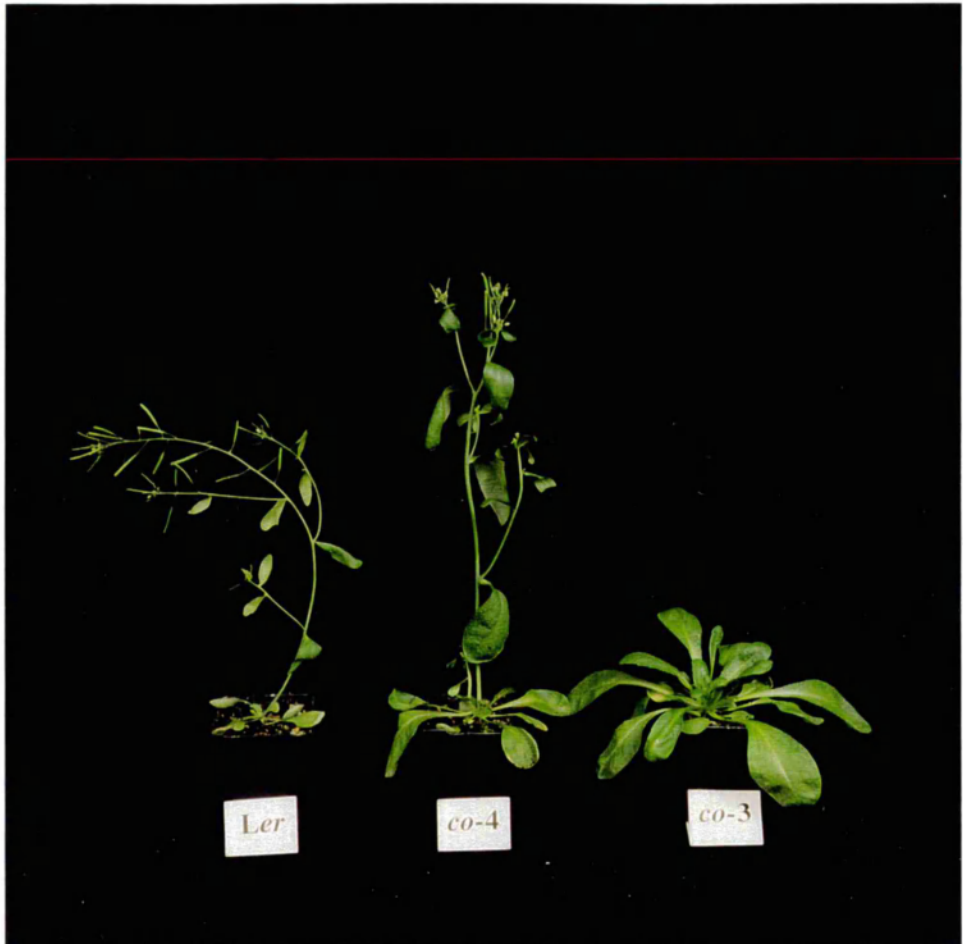


Figure 2.5 The *co* mutant plants carrying the weak *co-4* allele (centre) and the strong *co-3* allele (right) compared to wild type Landsberg *erecta* (left). All plants were grown in LDs and are 35 days old.

Subsequently Koornneef *et al.* (1991) described *co* as day-neutral; the LD- and SD-grown mutant flowered at the same time as wild type in SDs. Putterill *et al.* (1995) described *co-2* mutants as flowering later than wild type under LDs, and slightly earlier than wild type in SDs, as Rédei suggested.

To verify whether this was an allele specific effect seen only in *co-1* and *co-2*, or a characteristic of all *co* mutant alleles isolated so far, mutants carrying each of the seven *co* alleles were grown in a SD controlled environment room (CER; 10 h light, 14 h dark) alongside wild type Landsberg *erecta* control plants. All of the mutants flowered slightly earlier than wild type and with fewer leaves, although detailed leaf counts were not performed for all alleles. The differences between the alleles in flowering time and leaf number were not apparent in SDs. Figure 2.6 shows the distribution of flowering times within a population of *co-2* mutants and wild type Landsberg *erecta* plants grown in both LDs and SDs.

2.2.3 Other phenotypes

The *co* mutants growing in long-day inductive conditions resemble short-day grown *Arabidopsis* plants in their longer vegetative phase. They also exhibit a number of other phenotypes enhanced in, but not restricted to, short-day grown plants.

2.2.3.1 Chimeric flowering shoots

During the vegetative growth phase of *Arabidopsis*, leaf primordia are produced on the flanks of the apical meristem. In the axis of each leaf a bud forms, which develops into a secondary inflorescence or paraclade (Hempel and Feldman, 1995). The final secondary inflorescence to be initiated, immediately before the first fully formed flower on the primary shoot axis, is sometimes chimeric, being composed of part flower and part inflorescence (Hempel and Feldman, 1995). A secondary inflorescence whose development

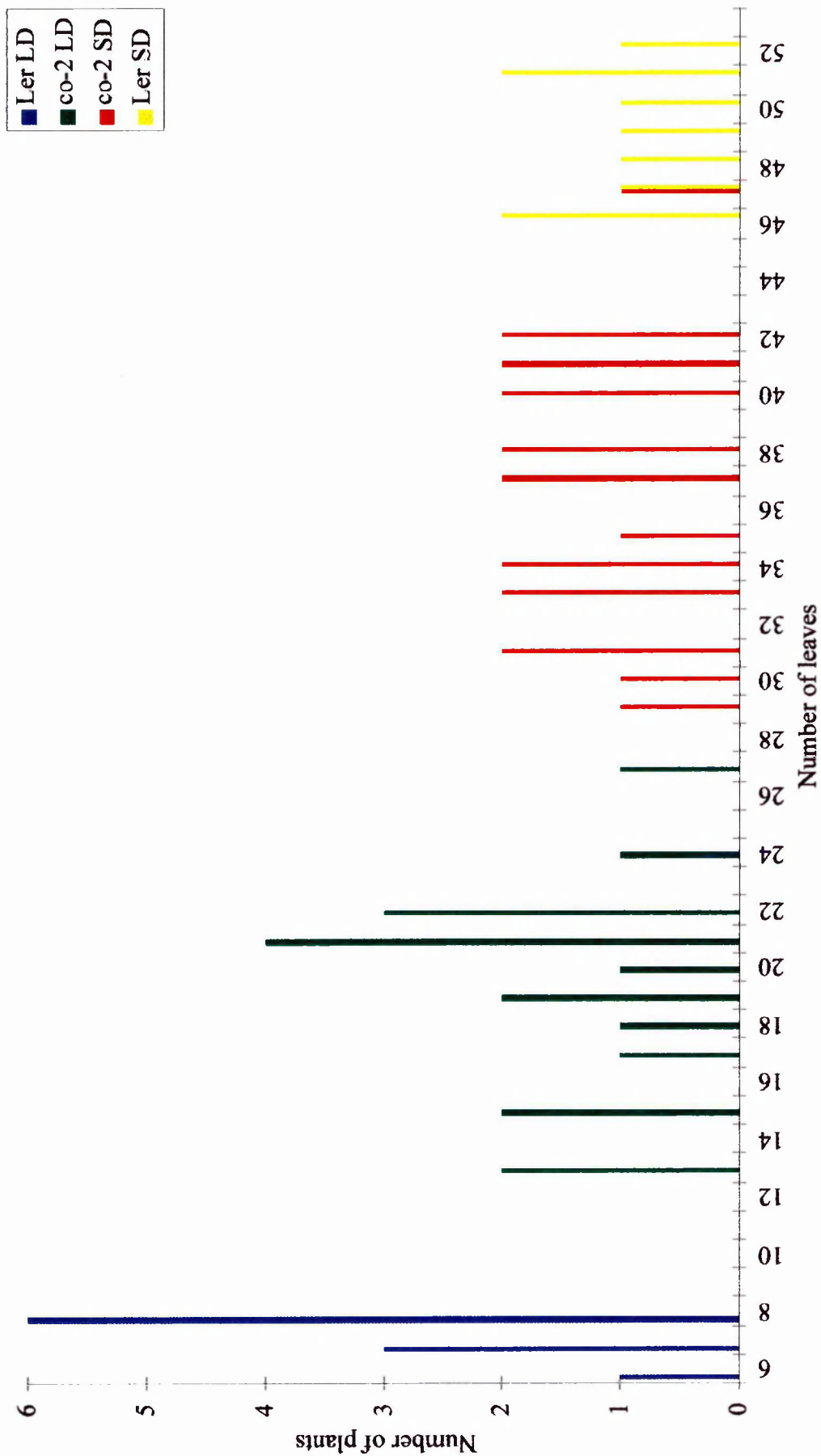


Figure 2.6 The flowering times, represented as total leaf number, of wild type *Landsberg erecta* plants and *co-2* mutant plants grown under both LD and SD conditions.

was initiated just prior to the transition from vegetative to reproductive growth can apparently be partially re-specified into a flower. This is thought to be a consequence of the floral stimulus acting directly on the primordium to specify floral identity. In SD non-inductive conditions, the frequency of plants carrying a chimeric shoot increases compared to LDs, possibly due to age-related differences in primordium sensitivity to a floral stimulus (Hempel and Feldman, 1995).

The frequency of chimeric shoots in LD-grown *co* mutants was also observed to be higher than in wild type under the same inductive conditions. The frequency of individuals with a chimeric shoot was recorded for some of the *co* alleles (Table 2.2, Figure 2.7). In general, the weakest alleles had the lowest frequency of chimeric shoots (*co-4*; 5%) and the more severe alleles, which more closely resemble short day grown wild type plants in their flowering time, had a higher frequency (*co-3*; 45%). Among the long day grown *co* mutants a gradation of phenotypes was observed, with some chimeric shoots being predominantly inflorescence-like or predominantly flower-like. This was also observed by Hempel and Feldman (1995) in wild type plants. A phenotype often seen in the *co* mutants was the emergence of a second flower from inside an altered carpel of the first flower-like part of the chimeric shoot. Whether this is also commonly seen among short day grown wild type plants is not known.

2.2.3.2 Altered number of floral organs

The first few flowers produced on the primary inflorescence of wild type plants often formed an increased number of petals and stamens when the plants were grown under SDs. This was also observed in LD-grown *co* mutants (Figure 2.8). Around 0-5 flowers on each mutant plant were altered, and these usually had between five and eight petals, and from seven to eleven stamens. The frequency of individuals with an altered number of floral organs in at least one flower, was recorded for some of the *co* mutants. The first ten

Mutant	Number of plants with altered flowers		Number of plants with a chimeric inflorescence	
	Number	%	Number	%
Ler	0/20	0%	1/20	5%
co-4	1/20	5%	1/20	5%
co-5	10/17	58%	1/17	6%
co-2	10/18	56%	4/18	22%
co-6	11/18	61%	7/18	39%
co-1	11/18	61%	8/18	44%
co-7	nt	nt	10/19	53%
co-3	nt	nt	9/20	45%

Table 2.2 Number of individuals with at least one flower with an altered number of petals and stamens, and the number of individuals with a chimeric inflorescence. All plants were grown in LDs.

A



B



C



Figure 2.7 Examples of chimeric shoots. **A)** A predominantly flower-like chimeric shoot with an altered gynoceium and a supernumary flower (SF) which also has an altered gynoceium. **B)** and **C)** Chimeric shoots that have a predominantly inflorescence-like phenotype. However, the shoots are partly organized as individual flowers. **S** = sepals, **P** = petals, **ST** = stamens. The flower and inflorescence components of **C** have been dissected apart for clarity. All examples are from *co-1* and *co-7* mutant plants grown in LDs. The shoots were all located below the lowest flower and above the uppermost secondary inflorescence on the primary shoot axis.

A**B****C**

Figure 2.8 Examples of flowers with an altered number of petals and/or stamens. **A)** A normal flower with four petals and six stamens. **B)** A flower with five petals and seven stamens. **C)** A flower with six petals. All flowers are from a *co-6* mutant plant grown in LDs.

flowers on the primary inflorescence of each individual were scored for the presence of this phenotype (Table 2.2). One individual in twenty plants homozygous for the weakest allele, *co-4*, exhibited the floral mutant phenotype. The frequency increased with the severity of the allele, and therefore the later the mutants flowered the more severe the floral phenotype. Mutants carrying all *co* alleles showed this phenotype, but it was most apparent in the intermediate alleles *co-5*, *co-2*, *co-6* and *co-1*. The first few flowers of the very late alleles *co-7* and *co-3* in this analysis did not open properly, possibly due to environmental effects, so the number of individuals carrying altered flowers was not recorded for these two alleles. An increase in floral organ number was not observed in any of the LD-grown Landsberg *erecta* control plants (Table 2.2) although this can happen at a low frequency (G. Coupland, personal communication).

2.3 The semi-dominance of the *constans* mutations

Rédei (1962) found the *co-1* mutation to be recessive when grown in continuous light, which were the conditions he used for his long day analyses. However, when this mutant and the six EMS-induced mutants were analyzed by Koornneef *et al.* (1991) in long day conditions of 16 h light, 8 h dark, it became apparent that they were all semi-dominant. The flowering time and leaf number of heterozygotes was intermediate between those of wild type and *co* homozygotes.

In order to confirm the semi-dominance of *co-1*, *co-4* and *co-3* demonstrated by Koornneef *et al.*, (1991) and to extend the analysis to all seven alleles, they were each independently crossed to Landsberg *erecta* and the F₁ progeny sown in LDs, along with the parental wild type and homozygous mutant controls. The flowering time and leaf numbers were scored as previously described (Table 2.3, Figures 2.9 and 2.10). All seven alleles were found to be semi-dominant, with intermediate flowering times and leaf numbers. This was confirmed in the F₂ generation, which produced the expected range of flowering times.

Mutant	Days to induction of flowering		Total leaf number	
	Heterozygote	Homozygote	Heterozygote	Homozygote
Ler	-	17.8±0.3	-	6.8±0.3
co-4	21.2±0.3	29.5±0.7	9.6±0.2	12.3±0.6
co-5	24.5±0.6	35.8±0.8	11.5±0.4	17.0±0.6
co-2	28.7±0.5	36.7±0.8	13.3±0.3	20.9±0.9
co-6	24.7±0.5	35.5±0.9	11.9±0.3	21.0±0.9
co-1	27.5±0.6	35.5±0.8	13.5±0.3	20.8±1.3
co-7	22.6±0.4	35.0±1.1	10.2±0.2	21.0±0.6
co-3	30.4±0.7	38.2±0.9	15.2±0.6	24.9±0.9

Table 2.3 Flowering times and leaf numbers in LDs of F₁ progeny of crosses between each *co* mutant and Landsberg *erecta*, compared to the homozygous parental mutants and Landsberg *erecta* .

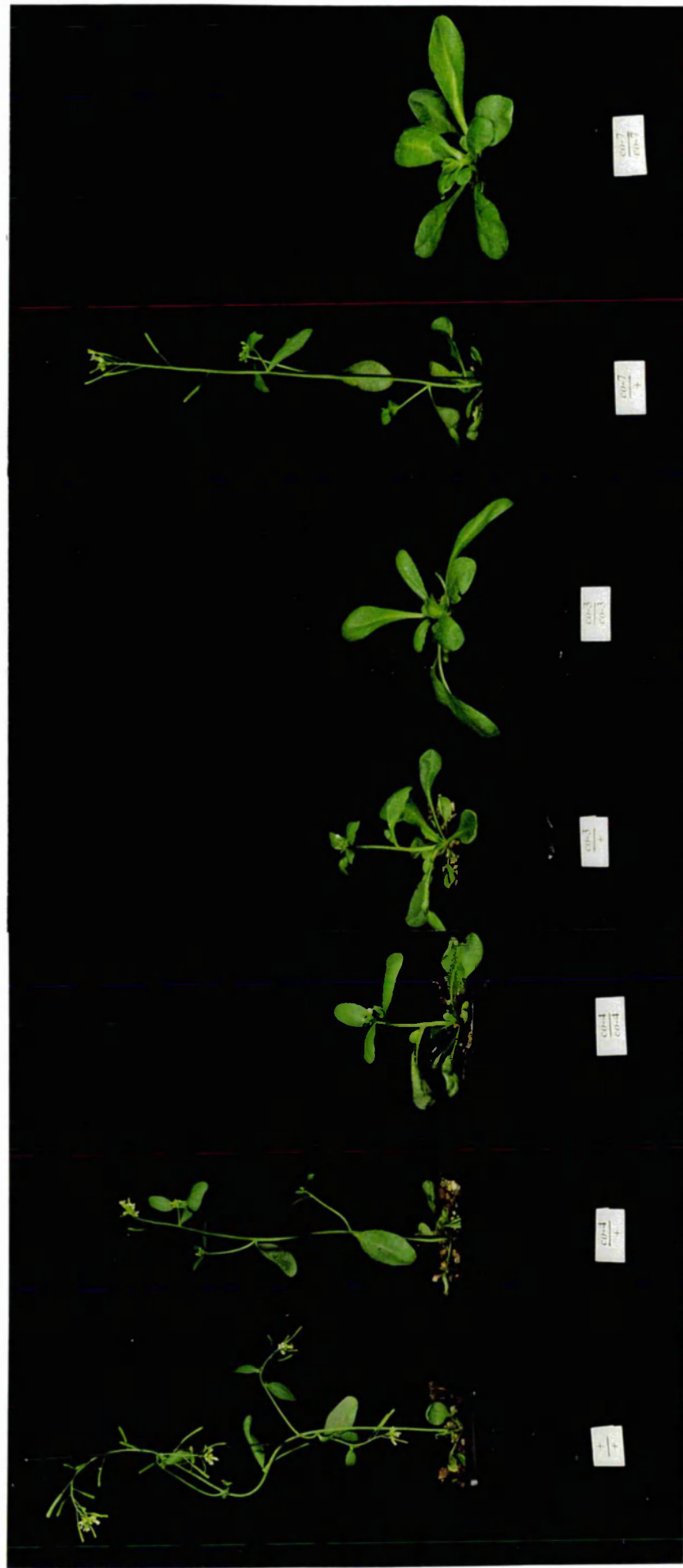


Figure 2.9 From the left, a wild type *Landsberg erecta* plant, a *co-4* heterozygote, a *co-4* homozygote, a *co-3* heterozygote, a *co-3* homozygote, a *co-7* heterozygote and a *co-7* homozygote. All plants were grown in LDs and are 28 days old.

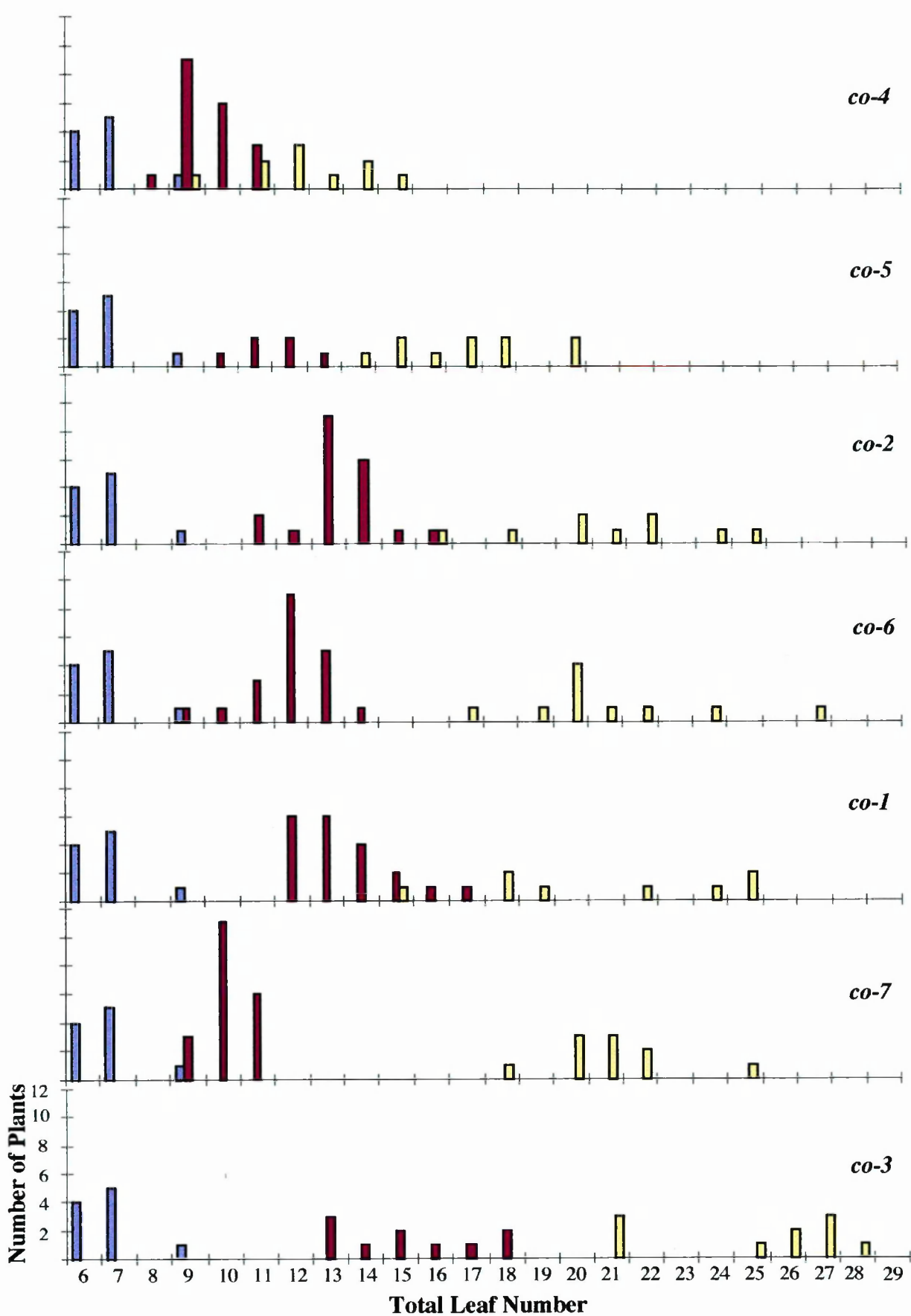


Figure 2.10 Bar charts showing the distribution of leaf numbers in the seven *co* mutant populations compared to the heterozygotes and wild type *Landsberg erecta*. X and Y axis labels are only shown on one chart for simplicity. The scale is the same for all charts.

Landsberg erecta plants are shown in blue
The heterozygous *co* mutants are in red
The homozygous *co* mutants are in yellow

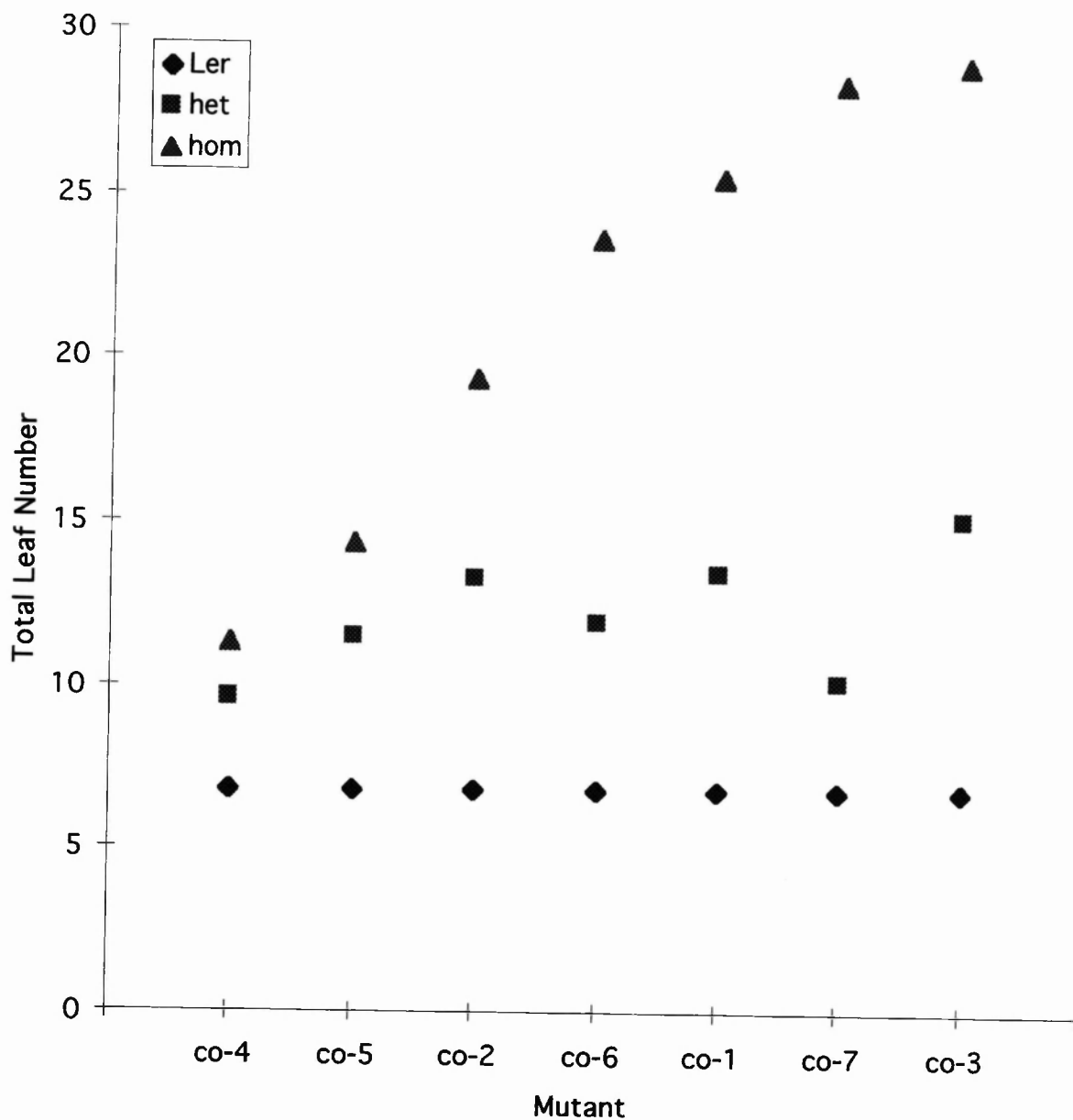


Figure 2.11 Summary of leaf numbers of the seven homozygous *co* mutants in LDs compared to the heterozygotes and wild type *Landsberg erecta*. Error bars are not included as they are smaller than the symbols.

Although all seven alleles are semidominant, the severity of the heterozygous phenotype is not directly related to the phenotype of the homozygotes (Figure 2.11). The heterozygotes for the weaker alleles have flowering times nearer to the homozygous mutants than to wild type, while for the stronger alleles the heterozygotes have flowering times closer to wild type than to the homozygous mutants. This is particularly apparent for the severe *co-7* allele (Figure 2.9, Figure 2.11) which is almost completely recessive.

2.4 Discussion

The flowering times of mutants carrying each of the seven *co* alleles (Rédei, 1962; Koornneef *et al.* 1991) were characterized under LDs and SDs and the alleles ranked in order of severity. The description of phenotype was thought likely to assist in interpreting the molecular analysis of each of the *co* alleles (Chapter 3). Each of the mutant alleles delayed the onset of flowering in LDs but in SDs flowering occurred at approximately the same time as (but very slightly earlier than) wild type. This had previously led to the *CO* gene being assigned to the LD promotion pathway (Koornneef *et al.* 1991), that was proposed to promote flowering specifically in response to long days (described in Chapter 1).

The results presented in this chapter are broadly in agreement with those of Koornneef *et al.* (1991). The main difference was the parameter used to measure flowering time. Koornneef *et al.* used the number of days from sowing to the opening of the first flower, while the first appearance of buds in the centre of the rosette was used here. In this analysis the order of severity of the alleles was the same whether total leaf number or flowering time was measured whereas in the Koornneef analysis the order was different. This was probably partly due to environmental effects on the opening of flowers, a point which was discussed in Koornneef *et al.* (1991), where it was suggested that the total leaf number might be a more accurate method for measurement of flowering time. The non-

genetic effects upon the opening of flowers was also noticed in this analysis; in a number of experiments it was the most variable parameter measured. The order of opening of flowers differed from the order of first appearance of buds and total leaf number, and was also different from that recorded by Koornneef *et al.*, suggesting that as a measure of flowering time it is unreliable.

The mutant alleles delayed flowering to different extents in LDs but all of the mutants flowered at approximately the same time in SDs. The quantitative differences in severity between the alleles suggests that a range of different types of mutation is present at the *CO* locus, and that both mutations with a weak effect on CO function as well as those with a more detrimental effect can be distinguished from wild type. Rédei (1962) and Koornneef *et al* (1991) originally isolated ten *co* alleles: one X-ray induced allele (Redei, 1962; *co-1*) and nine from an EMS mutagenesis (Koornneef *et al.*, 1991; *co-2* to *co-7*, plus three others which are no longer available). This is a relatively large number compared to some of the other flowering time loci, as would be expected for a gene with a strong effect on flowering time. Multiple alleles were also isolated at the *gi* and *fca* loci, and some of these are among the latest flowering time mutants isolated. Mutants with a relatively weak effect on flowering time, such as *fd*, *fy* and *fe* are represented by only one allele, which might suggest that only strong or null alleles were recovered at these loci because leaky mutations are indistinguishable from wild type. Alternatively, if strong or null alleles at these loci are lethal, or have a different phenotype then perhaps only weak alleles were recovered.

The alleles at the *gi*, *fca* and *fve* loci also differ quantitatively from each other. Recently, the *FCA* gene was cloned and a number of mutant alleles characterized. The strongest, *fca-1*, (Macknight *et al.*, 1997) is an EMS-induced allele in which an in-frame stop codon was introduced by a point mutation, resulting in the loss of approximately one half of the protein including a conserved glycine-rich region. This allele is assumed to be a null. However, the mutations in *fca-3* and *fca-4* (both induced by X-irradiation) have also been

characterized; *fca-3* contains a point mutation and *fca-4* a large inversion. Both of these mutations are predicted to cause premature termination earlier in the protein than the premature termination caused by the *fca-1* mutation. In both cases this results in the loss of most or all of one of the two putative RNA-binding domains assumed to be important for FCA function, in addition to the conserved glycine-rich domain at the C-terminus (R. Macknight, unpublished results). However, these two alleles have a much weaker effect on flowering time than *fca-1*, so in this case it is difficult to correlate the effect of the mutation on the structure of the protein with the observed phenotype. Similarly, the flowering time mutant alleles at the *fca*, *fha* and *gi* loci which were isolated after X-irradiation or fast neutrons, and may therefore contain larger rearrangements, are not among the most severe alleles at the loci. This is also the case for the *co-1* allele which was induced by X-rays and is weaker than *co-7* and *co-3* which were both EMS induced (Koornneef *et al.*, 1991).

Although they differ quantitatively with respect to flowering time, the *co* alleles cannot easily be grouped into different categories with respect to phenotype, unlike allelic variation at other loci. For instance the *cop1* alleles were grouped into three distinct phenotypic classes: weak, strong and lethal (McNellis *et al.*, 1994), or the *trp1* alleles which were categorised as either tryptophan auxotrophs or prototrophs, with dramatically different morphologies (Rose *et al.*, 1996). Late flowering is a quantitative phenotype and the phenotypes caused by the *co* alleles showed considerable overlap. Furthermore, all of the *co* mutants are viable, eventually flower and are fertile. This may be a consequence of other flowering pathways which operate in the absence of a functional *CO* gene, or that *CO* function can be compensated for within the long day pathway. Alternatively it is possible that all alleles isolated so far at the *CO* locus are leaky and that a null allele would have a more severe effect on flowering.

The *co* alleles are all semi-dominant, with the heterozygotes having flowering times and leaf numbers intermediate between wild type and the homozygous *co* parent. During the complementation of the *co-2* mutant phenotype with cosmids containing the wild type *CO* gene, *co-2* transgenic plants containing one copy of the T-DNA with the wild type *CO* gene showed a flowering time phenotype intermediate between *co-2* mutants and wild type. Transformants carrying two copies of wild type *CO* in the *co-2* mutant background flowered at the same time as wild type (Putterill *et al.*, 1995). This suggested that the intermediate phenotype observed in the *co-2* heterozygote was as a consequence of haplo-insufficiency rather than a gain-of-function mutation altering the CO protein and causing it to interfere with the wild type protein function (Park and Horvitz, 1986). That the semi-dominance of the *co* alleles is in part caused by haplo-insufficiency is also consistent with the very low level of expression of the *CO* gene (Putterill *et al.*, 1995). The reduction in abundance of the functional transcript by 50% in a heterozygote might be sufficient to reduce *CO* activity below a threshold necessary for wild type function. At least *co-2* is therefore predicted to be a loss-of-function mutation rather than a dominant gain-of-function mutation. However, the relative severity of the alleles was different in homozygotes and heterozygotes. The phenotype of a heterozygote for a leaky mutation, containing one fully functional copy of the gene and one partly functional copy is expected to have a phenotype closer to wild type than a heterozygote of a stronger or null allele (Park and Horvitz, 1986). This did not hold for all of the *co* mutant alleles; for example *co-7* caused a strong homozygous phenotype but a weak phenotype as a heterozygote (Figure 2.11). It is possible, therefore, that in addition to haplo-insufficiency some of the alleles show a dominant gain-of-function phenotype. The mutant protein may, for example, interfere with wild type protein activity and cause the heterozygote to have a stronger than expected phenotype. The later flowering alleles, such as the severe *co-7*, may encode proteins which have no CO activity in the homozygote, but which cannot interfere, either, with CO

function in a heterozygote. This might explain the near-recessive phenotype of a *co-7* heterozygote.

Other dominant late flowering mutants have been isolated. These include: *fwa* (Koornneef *et al.*, 1991) which is almost completely dominant; *lhy* which is caused by the ectopic expression of a gene implicated in the regulation of the circadian clock (Simon and Coupland, 1996) and *fhy*, which affects the signal transduction pathway of phytochrome A (Johnson *et al.*, 1994). In addition, a number of genes which regulate the flowering time of different ecotypes of *Arabidopsis* have naturally occurring dominant alleles. These include *FLC* and *FRI*, dominant alleles of which control the lateness in a number of very late flowering ecotypes (Koornneef and Peeters, 1997). The semi-dominant *gai* (*gibberellic acid-insensitive*) mutation is caused by a dominant gain-of-function mutation. A number of radiation-induced derivative alleles of *gai* were isolated and found to have a phenotype indistinguishable from wild type (Peng *et al.*, 1997). Sequence analysis of *GAI*, *gai* and the new alleles confirmed that the derivative alleles were intragenic suppressors that inactivated the gene.

The combination of phenotypic analysis of the alleles and characterization at the molecular level of the mutations causing these phenotypes might contribute towards an understanding of the functional significance of the different regions within the CO protein and the nature of the semi-dominant phenotype. Chapter 3 describes the identification and characterization of the lesions in the *CONSTANS* gene in each of the seven alleles and the analysis of domains of the protein thought to be important for *CONSTANS* function.

3 Molecular characterization of the *constans* alleles

3.1 Introduction

A series of *constans* alleles was characterized with respect to their flowering time phenotype, and the alleles ranked in order of severity (Koornneef *et al.*, 1991; Chapter 2). Once the *CONSTANS* gene had been cloned, it was possible to correlate these phenotypes with alterations in the *CO* nucleotide sequence. Investigating the molecular nature of various mutations in a single gene can provide insights into the function of the wild type gene. For example, individual amino acid residues important for protein function may be identified, and may provide insights into the roles of particular domains. Also mutations which lead to a disruption of transcription may define regions that control gene expression. Furthermore, molecular analysis of an allelic series may also explain how different mutations differ in the severity of their effects upon gene function. Unambiguous null alleles may also be identified and allow the effect of the complete loss of protein function on phenotype to be assessed. For these reasons the nucleotide sequences of all seven *co* mutant alleles were determined.

3.2 Identifying the mutations in the seven *constans* alleles

3.2.1 Southern analysis

Once the *CONSTANS* gene had been located to a 48 kb region covered by three overlapping cosmids the analysis of the alleles at a molecular level was initiated. The first approach in the molecular analysis was to establish whether any of the mutations could be detected by restriction enzyme digestion of plant DNA, Southern blotting and hybridization

to probes containing the *CO* gene. DNA from wild type and each of the seven *co* mutants was extracted and digested using a number of restriction enzymes. A total of fourteen enzymes with either 6 base pair recognition sites or 4 base pair recognition sites were used to optimize the detection of large rearrangements, small deletions or point mutations. The DNA was fractionated on duplicate agarose gels, Southern blotted and probed independently with cosmids 129 and 142 (Figure 3.1b). Together these cosmids cover a region of approximately 30 kb and both successfully complemented the *co* mutant phenotype (Putterill *et al.*, 1995). Examples of the autoradiographs obtained are shown in Figure 3.1a. No obvious DNA rearrangements in any of the alleles were detected by this method. This was not surprising because at least for the six EMS-induced alleles which might be expected to carry point mutations the mutation would have had to occur within the bases encoding a restriction enzyme recognition site. However, the analysis demonstrated that the single radiation-induced allele (*co-1*) was not a large deletion or rearrangement.

3.2.2 PCR amplification and cloning of the *constans* alleles

The *CO* gene was located to a 6.4 kb region where the two cosmids used successfully for complementation, 129 and 142, overlap. This region was sequenced and two ORFs separated by 233 bp were identified (Putterill *et al.*, 1995). Primers flanking the ORFs were designed by Dr. J. Putterill (Figure 3.3) and PCR amplification using wild type DNA yielded a single 1.95 kb fragment which was cleaved by *EcoR* I into two fragments of 1.09 and 0.86 kb. This was in agreement with nucleotide sequence which identified a single *EcoR* I site within the proposed 233 bp intron. The PCR was repeated for all seven alleles. Duplicate amplification reactions were performed to identify and discount errors introduced by *Taq* polymerase. PCR products of the predicted size were amplified from all of the alleles. No changes were detectable at this stage. The 1.95 kb products were blunt ended using T4 DNA polymerase, cloned into the *EcoR* V site of pBluescript (SK+) and

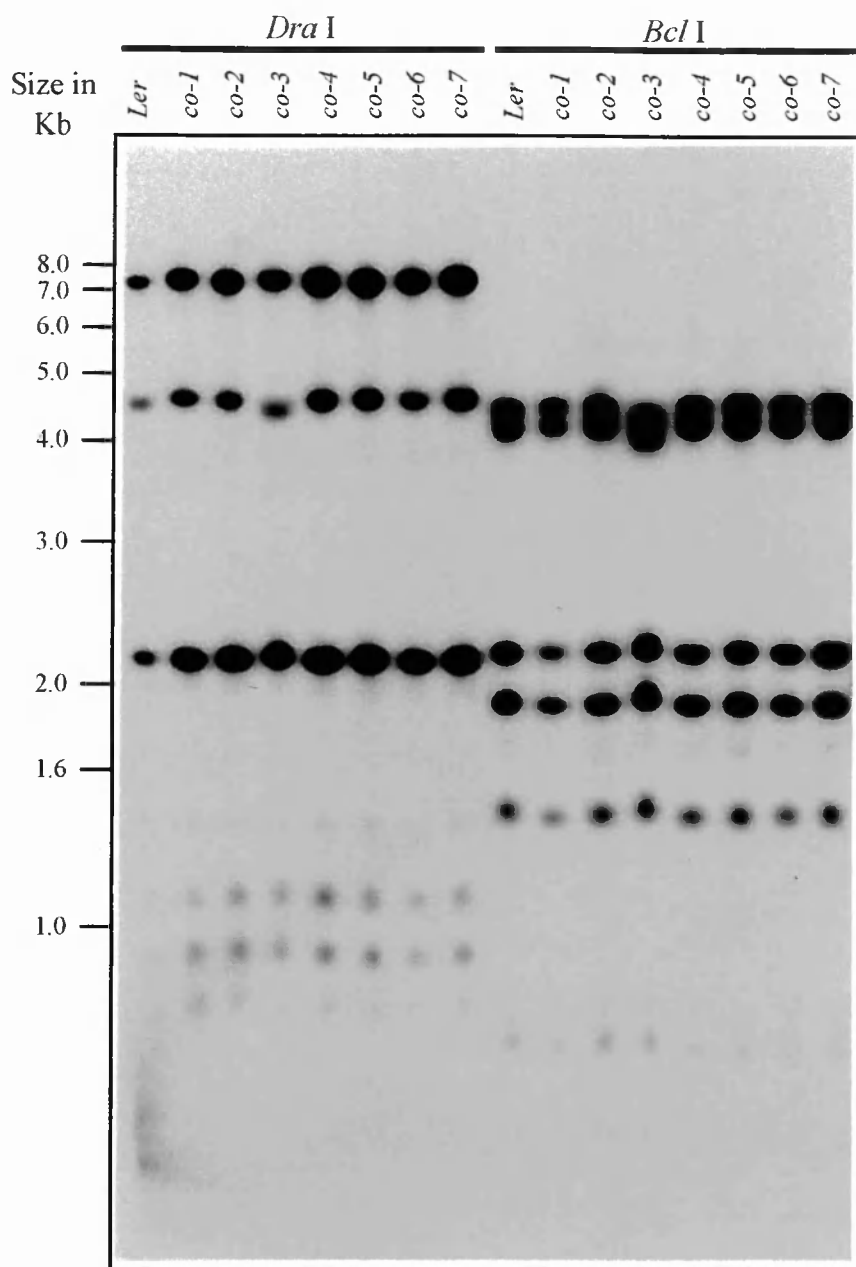


Figure 3.1 a Southern blot of plant DNA extracted from wild type Landsberg *erecta* and the seven *co* mutants , digested with the enzymes indicated and probed using cosmid 129 (see below).

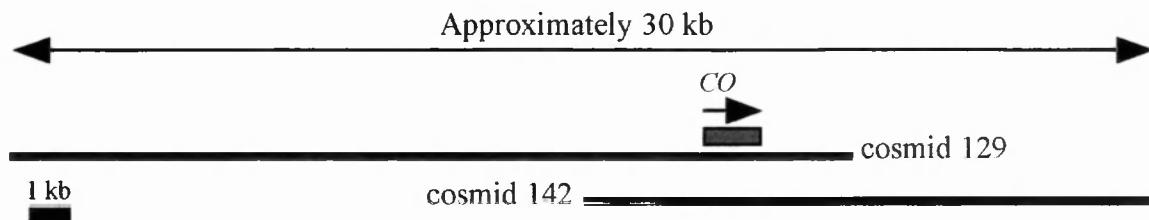


Figure 3.1 b The two cosmids which were used as probes against Southern blots of restriction digested DNA from the seven *co* mutants and wild type.

transformed into *E. coli* strain DH5 α . Ampicillin resistant colonies containing the correct recombinant plasmids were identified by digesting prepared plasmid DNA with *EcoR* I, which yielded a 0.86 or 1.09 kb fragment, depending on the orientation of cloning, and a larger fragment containing the vector.

3.2.3 Sequencing the *constans* alleles

Oligonucleotide primers had previously been designed and used to sequence the 6.4 kb overlap between the complementing cosmids to try to identify the *CO* gene. Sequencing of the alleles was initiated at the 5' end of each cloned ORF using these primers. As the sequencing progressed more primers were designed and made so that the entire coding sequence for both exons, the intron and the 5' and 3' untranslated regions could be analyzed. The first methionine of the CO protein had previously been identified by the presence of an in-frame stop codon three codons upstream of the candidate ATG. The 3' end of the transcript had also been identified by 3' rapid amplification of cDNA ends (RACE), and was located 160 bp downstream of the final stop codon. (Putterill *et al.*, 1995). In total, 1750 bp of the *CO* gene were sequenced from each allele, including 250 bp of 5' and 150 bp of 3' untranslated sequence (Figure 3.3). Two independent clones for each allele, amplified in different PCR reactions, were sequenced and compared with the ORFs cloned from wild type DNA. The sequencing reactions for a particular primer were arranged on sequencing gels in blocks of Gs, As, Ts and Cs to make identification of mutations easier. Wild type reactions were arranged in a normal GATC ladder so that sequence coverage of the gene could be monitored (Figures 3.2a and b).

Mutations in alleles *co-1*, *co-2* and *co-3* were identified first. This confirmed that the putative ORFs identified in the 6.4 kb complementing region were part of the *CONSTANS* gene. The mutations in alleles *co-1* and *co-2* were published in Putterill *et al.*, 1995 and Dr. J. Putterill identified the mutations in *co-1*, *co-2* and *co-3*. The mutations in *co-4*, *co-5*,

co-6 and *co-7* were identified by the author. Autoradiographs of sequencing gels revealing the mutations in these four alleles are shown in Figures 3.2a and b.

Figure 3.3 shows the 1950 bp region amplified by PCR from each cloned *CO* gene, with the position and nature of the mutations in the alleles identified by sequencing. The sequence of both exons and the single intron were determined for each allele to ensure that only one mutation had occurred in each allele and that these were therefore responsible for the mutant phenotypes. Other sequence changes found during the sequencing are marked. The *co* alleles are in a Landsberg *erecta* background but the cosmids used for complementation were from a library made from Columbia DNA (Olszewski *et al.*, 1988), and a number of polymorphisms between the two ecotypes were identified and are marked. Polymorphisms were detected in the untranslated 5' region of the gene, in the promoter and in the intron, but not in either exon. PCR errors identified in individual clones were recorded but are not shown. Whenever possible clones carrying no PCR errors in the ORF were identified for subsequent experiments (Chapter 5).

3.2.4 The mutations in the seven *constans* alleles define domains of the protein that are important for CO function

The mutations identified in the seven *co* alleles were found to be clustered in two regions of the predicted protein sequence. The six EMS alleles each contain a single point mutation, and each of them results in a single amino acid substitution. All mutations are G to A or C to T transitions, which is expected for an alkylating agent such as EMS (Koornneef *et al.*, 1982). The X-ray induced allele, *co-1*, first identified by Rédei (1962), contains a nine base pair deletion which precisely removes three amino acids. All seven mutations are predicted to be missense mutations within the protein. No obvious null mutations, or mutations affecting regulatory sequences of the *CONSTANS* gene were

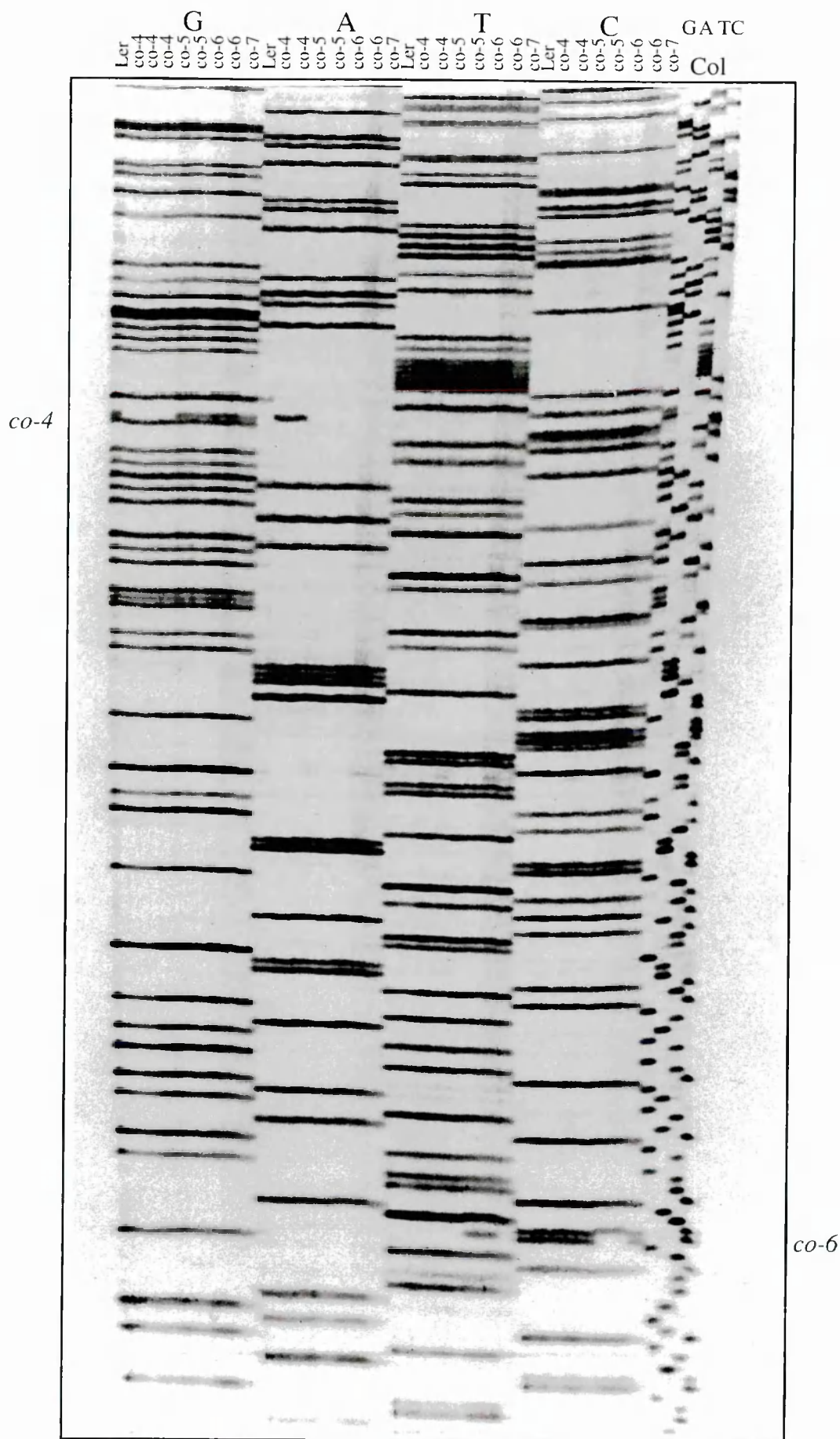


Figure 3.2a The sequencing gel that first identified the mutations in *co-4* and *co-6*. The *CONSTANS* gene was originally sequenced from cosmids from a library made from the Columbia (Col) ecotype so this was included to identify any polymorphisms between Landsberg *erecta* and Columbia. The primer used was co33 (Materials and Methods).

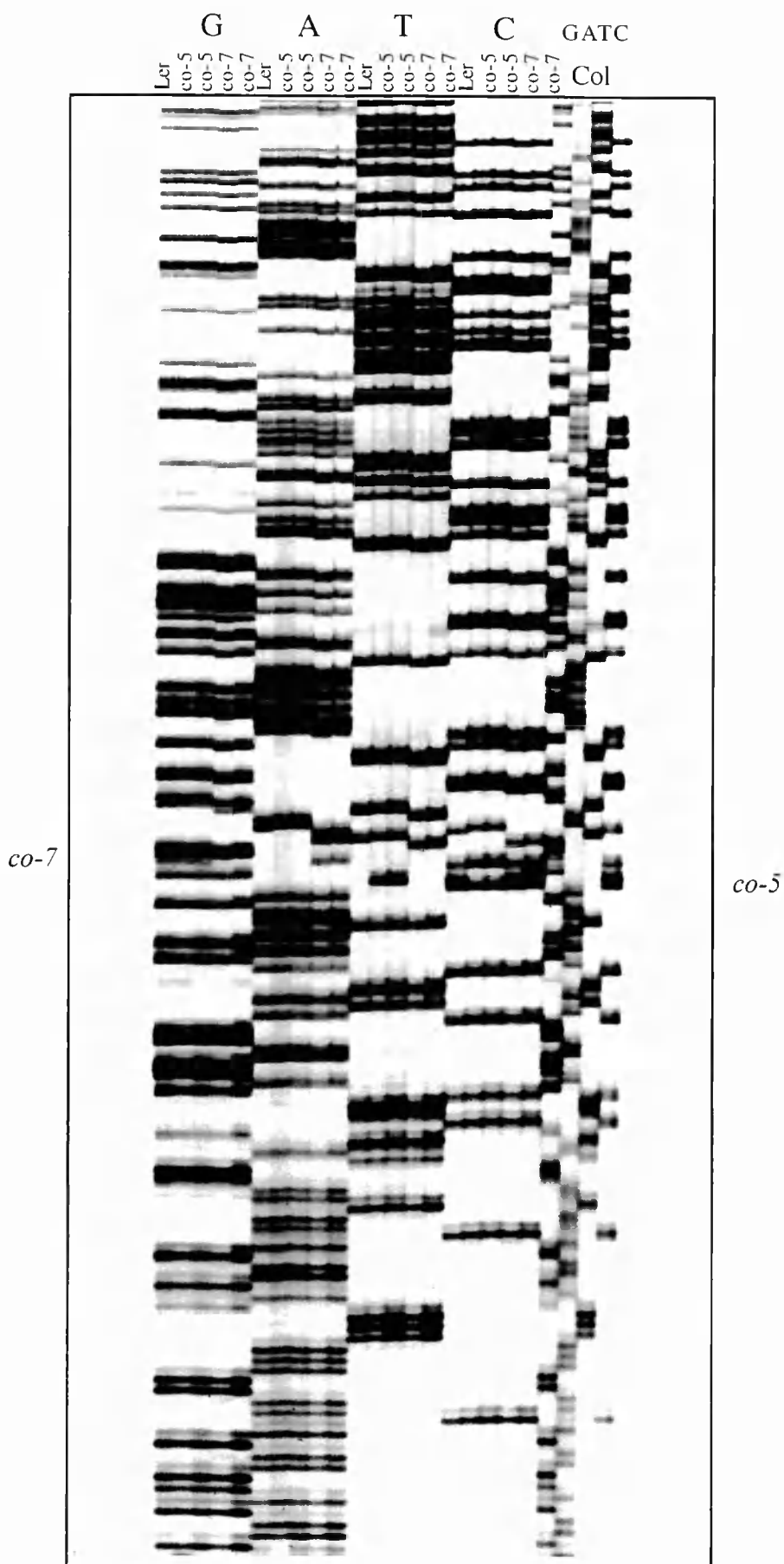


Figure 3.2b The sequencing gel that first identified the mutations in *co-5* and *co-7*. The primer used was co54 (Materials and Methods)

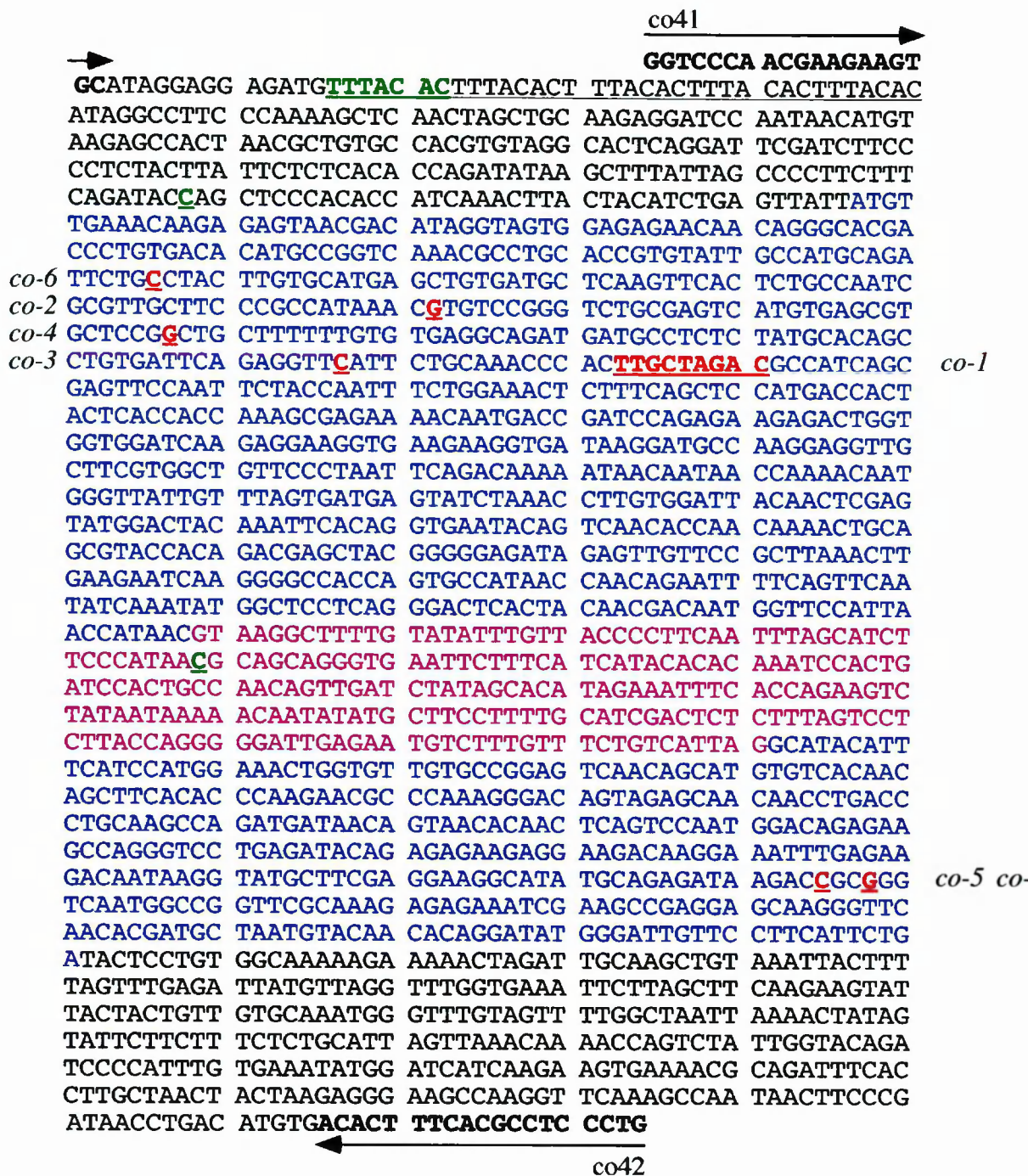


Figure 3.3 The *CO* genomic region amplified by PCR and sequenced from all the *co* mutants. The primers used for amplification, co41 and co42, are in bold. The two exons are in blue separated by the 233 bp intron in purple. The bases affected in each mutant allele are shown underlined and in red. Each of the single base substitutions in six of the alleles is either a C-T or G-A transition. The nine bases deleted in *co-1* are also underlined and in red.

The three polymorphisms detected between Landsberg *erecta* and Columbia ecotypes are shown in green. In all cases the Columbia sequence is shown. For the two single base changes (one in the 5' untranslated region and one in the intron), at the equivalent position in Landsberg *erecta* the C is replaced by an A. The seven base pair motif in the promoter (underlined and in green) is repeated five times in Columbia (underlined in black) but only four times in Landsberg *erecta*.

identified. Figure 3.4 shows the DNA sequence of *CO*, the predicted protein sequence and the changes in each of the mutant alleles.

co-1; The *co-1* mutant was originally isolated as one of four "supervital" mutants from an X-irradiation screen by Rédei (1962). The mutation is a nine base pair deletion (TTGCTAGAC; bases 287 to 295 - the A residue of the first ATG codon designated as base 1) which precisely removes three amino acids; leucine (at amino acid position 96), alanine 97 and arginine 98 (Figure 3.4). The sequences flanking the deletion reconstruct the C-GC codon of arginine 99. The three deleted amino acids are located in the basic region of the putative DNA binding domain, at the carboxyl side of zinc finger 2. The secondary structure prediction program PHD predict protein (Rost 1996) predicts a short helix in CONSTANS, over a region of six amino acids (PLARRH, Figure 3.5a) including the LAR motif deleted in *co-1*.

The remaining six *co* mutants were isolated by EMS mutagenesis and screening for late flowering mutants (Koornneef *et al.*, 1991). Four of these also have mutations which affect the zinc finger domain. All sequence changes are represented in Figure 3.4.

co-2; The *co-2* allele carries a mutation which affects the conserved, basic domain on the carboxyl side of zinc finger 1. A G to A transition at position 176 of the DNA sequence substitutes histidine for arginine 59.

co-3; The *co-3* allele caused the strongest late flowering phenotype (Koornneef *et al.*, 1991; Chapter 2). A single base change leading to a single amino acid substitution is responsible for this severe phenotype. At DNA sequence position 271 a C to T transition substitutes tyrosine for histidine 91. This histidine is located in the basic domain on the carboxyl side of zinc finger 2.

co-4; This allele caused the weakest late flowering phenotype (Koornneef *et al.* 1991; Chapter 2). The lesion in *co-4* was found in zinc finger 2 at DNA position 211, where a G to A transition substitutes threonine for alanine 71. The region from proline 70 to alanine 77, including alanine 71 affected in this allele, is predicted in the wild type CO protein to form a helix. The mutation in *co-4* is predicted to shorten this helix and introduce a short stretch of β -sheet (Figure 3.5a).

co-6; The *co-6* mutation causes a C to T transition at position 110 of the DNA sequence and substitutes valine for alanine 37 within zinc finger 1. The alanine residue affected is located within two predicted beta sheets joined by a short loop. The mutation is predicted to lengthen the second beta sheet and shorten the intervening loop (Figure 3.5a).

Two *constans* alleles, *co-5* and *co-7*, that were induced with EMS contained mutations that affect a basic domain near the C-terminus of the protein (Figure 3.4). The residues affected in these alleles are found in the most highly conserved part of the basic domain (see section 3.3) and are directly adjacent to one another.

co-5; The *co-5* mutant is relatively weak, second only to *co-4* (Chapter 2). It also contains a point mutation; a C to T transition at position 1016 of the DNA sequence results in the substitution of leucine for proline 339. The secondary structure prediction program "PHD predict protein" predicted two alpha helices separated by a short loop in this basic domain (discussed later). The *co-5* mutation is predicted to shorten the intervening loop or turn (figure 3.5b).

co-7; The *co-7* mutant is the second strongest, after *co-3* (Chapter 2). At position 1019 a G to A transition substitutes glutamine for arginine 340. Again, like the *co-5* mutation, the effect of the substitution in *co-7* on the predicted secondary structure is to shorten the loop between two alpha helices.

ATGTTGAAACAAGAGAGTAACGACATAGGTAGTGGAGAGAAACAACAGGGCAGCACCTGT 60
 M L K Q E S N D I G S G E N N R A R P C
 GACACATGCCGGTCAAACGCCTGCACCGTGTATTGCCATGCAGATTCTGCTACTTGTGC 120 *co-6*
 D T C R S N A C T V Y C H A D S ^T**A** Y L C
 ATGAGCTGTGATGCTCAAGTTCACTCTGCCAATCGCGTTGCTTCCCGCCATAAACGTGTC 180 *co-2*
 M S C D A Q V H S A N R V A S R H K ^A**R** V
 CGGGTCTGCGAGTCATGTGAGCGTGCTCCGCTGCTTTTTTGTGTGAGGCAGATGATGCC 240 *co-4*
 R V C E S C E R A P ^A**A** F L C E A D D A
 TCTCTATGCACAGCCTGTGATTCAGAGGTTCTATTCTGCAAAACCCACTTTGCTAGACGCCAT 300 *co-3/co-1*
 S L C T A C D S E V ^T**H** S A N P L **A** R R H
 CAGCGAGTTCCAATTCTACCAATTTCTGGAAACTCTTTCAGCTCCATGACCACTACTCAC 360
 Q R V P I L P I S G N S F S S M T T T H
 CACCAAAGCGAGAAAACAATGACCGATCCAGAGAAGAGACTGGTGGTGGATCAAGAGGAA 420
 H Q S E K T M T D P E K R L V V D Q E E
 GGTGAAGAAGGTGATAAGGATGCCAAGGAGGTTGCTTCGTGGCTGTTCCCTAATTCAGAC 480
 G E E G D K D A K E V A S W L F P N S D
 AAAAATAACAATAACCAAAACAATGGGTTATTGTTTAGTGATGAGTATCTAAACCTTGTG 540
 K N N N N Q N N G L L F S D E Y L N L V
 GATTACAACCTCGAGTATGGACTACAAATTCACAGGTGAATACAGTCAACACCAACAAAAC 600
 D Y N S S M D Y K F T G E Y S Q H Q Q N
 TGCAGCGTACCACAGACGAGCTACGGGGGAGATAGAGTTGTTCCGCTTAAACTTGAAGAA 660
 C S V P Q T S Y G G D R V V P L K L E E
 TCAAGGGGCCACCAAGTCCATAACCAACAGAATTTTCAGTTCAATATCAATATGGCTCC 720
 S R G H Q C H N Q Q N F Q F N I K Y G S
 TCAGGGACTCACTACAACGACAATGGTTCATTAAACCATAACGCATACATTTTCATCCATG 780
 S G T H Y N D N G S I N H N A Y I S S M
 GAAACTGGTGTGTGCGGGAGTCAACAGCATGTGTCACAACAGCTTCACACCCAAGAACG 840
 E T G V V P E S T A C V T T A S H P R T
 CCCAAAGGGACAGTAGAGCAACAACCTGACCCTGCAAGCCAGATGATAACAGTAACACAA 900
 P K G T V E Q Q P D P A S Q M I T V T Q
 CTCAGTCCAATGGACAGAGAAGCCAGGGTCTGAGATACAGAGAGAAGAGGAAGACAAGG 960
 L S P M D R E A R V L R Y R E K R K T R
 AAATTTGAGAAGACAATAAGGTATGCTTCGAGGAAGGCATATGCAGAGATAAGACCGCGG 1020 *co-5/co-7*
 K F E K T I R Y A S R K A Y A E I R ^T**P** ^A**R**
 GTCAATGGCCGGTTCGCAAAGAGAGAAATCGAAGCCGAGGAGCAAGGGTTCAACACGATG 1080
 V N G R F A K R E I E A E E Q G F N T M
 CTAATGTACAACACAGGATATGGGATTGTTCTTCATTCTGA 1122
 L M Y N T G Y G I V P S F .

Figure 3.4 The DNA sequence of the *CO* ORF and the predicted amino acid sequence of the protein. The amino acids shown in bold are those within the predicted zinc finger region. The sequence changes in the seven mutant alleles are shown in red. For the point mutations the mutant sequence is shown above the wild type DNA sequence and the predicted amino acid substitution is shown below the wild type amino acid sequence. The nine base pair deletion in *co-1* and the three amino acids predicted to be deleted are also shown in red.

A	Wild type Zinc Finger 1	CDTCRSNACTVYCHADSAYLCMSCDQVHSANRVASRHKVRV
	Predicted structureEEEE.....HHHHHH.....EEE
	Probability	7336678935865323226653431221141267532474576
B	Wild type Zinc Finger 2	CESCERAPAAFLCEADDASLCTACDSEVHSANPLARRHQRVPI
	Predicted structure	E.....HHHHHHHH..HHH.....HHHHHH.....E
	Probability	3142121799986535311124665322575634786378772
	Mutant Zinc Finger 2	CESCERAPTAF LCEADDASLCTACDSEVYSANPRHQRVPI
	Predicted structure	EE.....HHHHEE.....EE.....E
	Probability	6311366443232014533233455311142997583215

Figure 3.5a The predicted protein sequence of CONSTANS zinc finger 1 (Panel A) and zinc finger 2 (Panel B) with the prediction of secondary structure and the calculated probability of a secondary structure being formed underneath. The prediction is from the PHD predict protein program (Rost, 1996). For each zinc finger, both the wild type protein sequence and a mutant protein sequence incorporating the predicted changes in all the mutant alleles is shown, with the residues affected by the mutations in red. The wild type sequence was rerun through the program after the mutant sequences to check for variation between runs.

H = predicted helix

E = predicted β sheet (extended)

. (period) = predicted loop, turn or other.

Wild type C-terminal basic domain	LSPMDREARVLRYPEKTRKFEKTIYASRKAYAEIRPQVNGRFAKREIEAEEQGNTMLMYNTGYGIVPSF
Predicted structureHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH.....EEE.....EEEE..
Probability	4793279999998765415567888999999999996212043145789899973884489972597455489
Mutant C-terminal basic domain	LSPMDREARVLRYPEKTRKFEKTIYASRKAYAEIRLQVNGRFAKREIEAEEQGNTMLMYNTGYGIVPSF
Predicted structureHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH.....EEE.....EEEE..
Probability	47932799999986641556788899999999998335301477889999973884489972597455489

Figure 3.5b The predicted protein sequence of CONSTANS C-terminal basic domain with the prediction of secondary structure and the calculated probability of a secondary structure being formed underneath. The prediction is from the PHD predict protein program (Rost, 1996). The wild type protein sequence and a mutant protein sequence incorporating the predicted changes in the two mutant alleles is shown, with the residues affected by the mutations in red.
H = predicted helix
E = predicted β sheet (extended)
.(period) = predicted loop, turn or other.

The clustering of the mutations into two regions; *co-1*, *co-2*, *co-3*, *co-4* and *co-6* affecting the putative zinc fingers and *co-5* and *co-7* affecting a C-terminal highly positively charged domain (summarised in Table 3.1) and the conservation of these regions in a number of homologues (discussed in section 3.3) suggests that they define domains of the protein that are important for CONSTANS function.

The protein has a number of other domains characteristic of transcription factors that may be important for function. Two of these are addressed in detail in the discussion (section 3.4) but are mentioned here as they appear in several figures. The C-terminal basic domain includes a stretch of amino acids, from arginine 306 to lysine 321, that could serve as a nuclear localization signal. This region was not affected by either of the mutations in alleles *co-5* and *co-7* (Figure 3.6). In addition, there are five regions which potentially could serve as transcriptional activation domains, although the most likely candidate is an acidic domain which extends from glutamic acid 124 to glutamic acid 150. This region contains 11 acidic residues out of 27 and has a net negative charge of -6 (Figure 3.6). Another feature of acidic activation domains is their ability to form alpha helices, as demonstrated for the yeast transcription factor GAL4 (reviewed in Ptashne, 1988, Mitchell *et al.*, 1989). Using the PHD predict protein program, a helix was predicted in CO from residues 146 to 156, a short region which overlaps with the proposed activation domain.

3.2.5 Analysis of expression of *co* mutant alleles

In wild type *Arabidopsis* the *CONSTANS* mRNA is present at levels undetectable by conventional Northern analysis and RT-PCR was therefore used to analyze *CO* expression (Putterill *et al.*, 1995). Using this method a transcript indistinguishable from that of wild type was detected in the *co-2* mutant (J. Putterill, R. Simon, unpublished results), so a similar approach was taken with the remaining alleles. Total RNA was extracted from

Allele	Mutation	Domain affected	Total leaf number
<i>co-4</i>	alanine to threonine	Zinc fingers	11.3
<i>co-5</i>	proline to leucine	C-terminal basic	14.4
<i>co-2</i>	arginine to histidine	Zinc fingers	19.4
<i>co-6</i>	alanine to valine	Zinc fingers	23.7
<i>co-1</i>	in-frame deletion of leucine, alanine, arginine	Zinc fingers	25.6
<i>co-7</i>	arginine to glutamine	C-terminal basic	28.5
<i>co-3</i>	histidine to tyrosine	Zinc fingers	29.1

Table 3.1 The mutations in the seven *co* alleles, the domains of the CO protein they affect and the mean flowering times of the *co* mutants represented as total leaf numbers.

duplicate samples of 10 day old wild type and mutant seedlings. At this stage the first two leaves were expanding and the *CO* mRNA was readily detectable by RT-PCR in wild type plants (R. Simon, unpublished results). RT-PCR was performed using primers co49 and Oli9 (Chapter 7, Materials and Methods) which amplify a fragment at the 5' end of the *CO* cDNA. Oli9 was designed by R. Simon to exon sequence at either side of the single intron and therefore to specifically amplify only *CO* cDNA and not contaminating genomic DNA (Figure 3.7b). The products of the PCR reactions were fractionated on an agarose gel, Southern blotted and probed with a *CO*-specific probe. The results of this RT-PCR experiment are not quantitative. However, in all cases a hybridising fragment of the expected size for a wild type transcript was amplified from all seven of the previously isolated mutant alleles and the wild type controls (Figure 3.7a). No hybridizing fragment was detected in a control sample lacking cDNA. Allele *co-8*, was included in this experiment (Chapter 4). As the *co-8* allele contains a deletion at the 5' end of the gene which removes one of the primer annealing sites, this served as a negative control. Alleles *co-1* to *co-7* all contained a transcript that was detectable by RT-PCR.

3.3 Homology to other genes

Sequencing of the region around the *CONSTANS* gene identified a closely related homologue, *CONSTANS-LIKE 1* (*COL1*, GenBank accession numbers Y10555/Y10556), which was found 3.6 kb proximal to *CO* and transcribed in the same direction (Putterill *et al.*, 1997). The predicted proteins are 67% identical overall, but over an 86 amino acid region corresponding to the putative zinc fingers are 86% identical. At the C-terminus of the protein the basic domain is also conserved between the two proteins, with 87% identity over a region of 47 amino acids (Figure 3.8). The high conservation of these two regions suggested that these regions were important for protein function and that *CO* and *COL1* might have related functions.

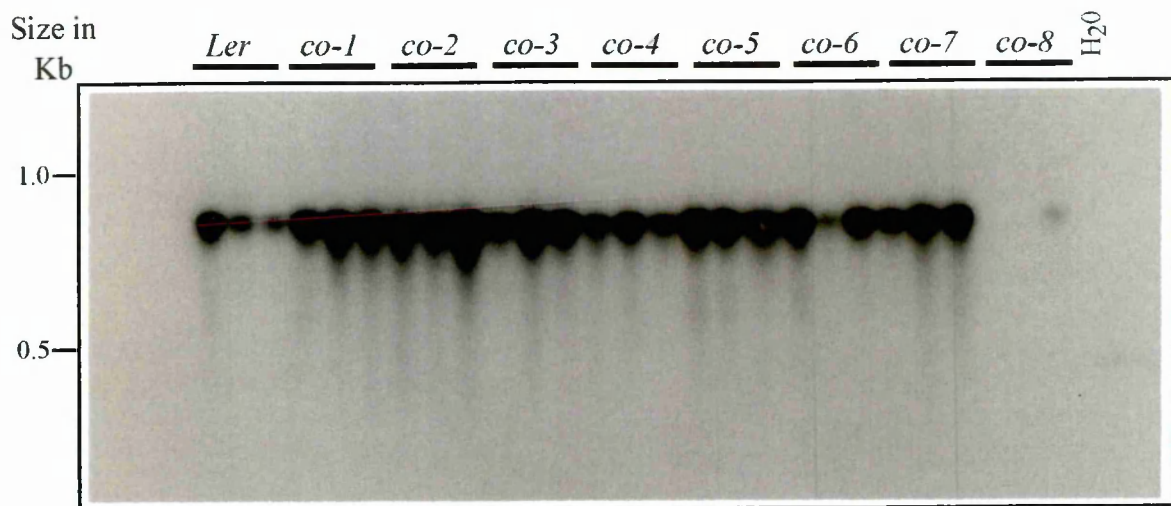


Figure 3.7a Southern blot of RT-PCR products from wild type Landsberg *erecta* and the *co* mutants. RNA was extracted from three duplicate batches of plant material for each mutant or wild type, cDNA prepared from each and PCR performed using *CO*-specific primers *co49* and *oli9*. The Southern blot was probed with the *CO* cDNA. A no-cDNA PCR control was loaded into a single lane on the right. The faint band present in sample 3 in mutant *co-8* is assumed to be PCR contamination. This mutant, which has a deletion which removes primer *co49* (Chapter 4) is the negative control.

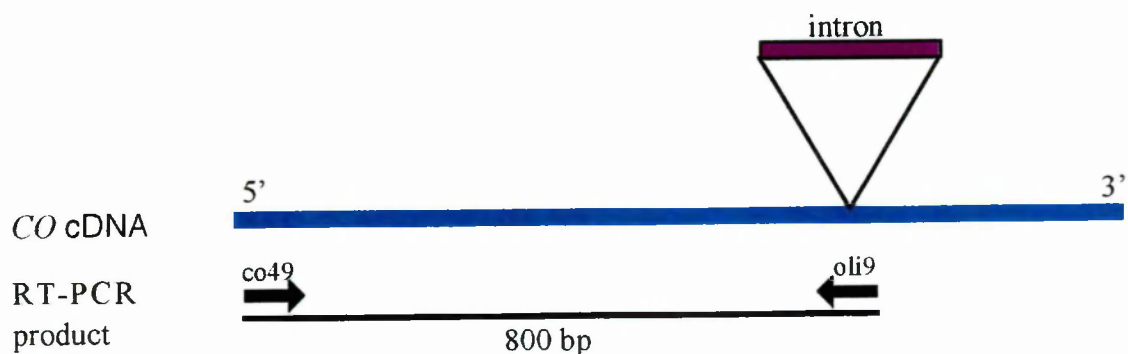


Figure 3.7b The PCR primers used to amplify the fragment, shown in 3.7a, from cDNA prepared from RNA extracted from each of the *co* mutants. *oli9* (Materials and Methods) was designed to exon sequence on either side of the intron in order to amplify only cDNA and not contaminating genomic DNA.

Arabidopsis genomic library screens using the *CONSTANS* gene as a probe show that a number of *CONSTANS*-like genes exist (M. Igeño, unpublished results). The gene *CONSTANS-LIKE 2* (*COL2*, accession numbers L81119/L81120) was isolated in this way (Ledger *et al.*, 1996). This homologue is 68% identical to the *CONSTANS* protein, with 84% and 90% identity over the zinc fingers and the C-terminal basic domain, respectively (Figure 3.8) The function of *COL1* and *COL2*, and whether or not they are involved in the regulation of flowering is not yet known.

Subsequently other *CO* homologues have been reported. These are discussed below and Figures 3.9, 3.10 and 3.11, and Tables 3.2 and 3.3 summarise some of the available data. The recently isolated *CO* orthologues from *Brassica napus* have not been included. These proteins are so highly conserved that it was felt that they would not contribute significantly to this analysis. Chapter 6 describes in detail the characterization of the *B. napus* orthologues and their conservation of function.

Recently, a cDNA which was able to complement the salt-sensitive phenotype of yeast calcineurin mutants was isolated from *Arabidopsis* (Lippuner *et al.* 1996). This was shown to be homologous to *CONSTANS* over the zinc finger domain, with 39% identity. There is a small basic domain at the carboxyl terminus of the protein but little other homology.

The predicted *CONSTANS* protein sequence was compared to sequences in the GenBank databases using the BLAST programs (Altschul *et al.*, 1990). The first searches found no significant homologies (Putterill *et al.*, 1995). However, recent searches have detected at least 30 homologues identified in DNA sequencing programmes of several species. Using the TBLASTN comparison to the GenBank databases, in addition to the published proteins described above, three homologues from the *Arabidopsis* genomic sequencing programme were identified. Z97338 is on the annotated fragment ATFCA3, chromosome 4, in the ESSA 1 region (European Scientists Sequencing Arabidopsis, Bevan *et al.*, 1998). It is homologous to *CO* over the zinc finger domain with 44% identity. AC002332 was found

Constans	MLKQESNDIG	SGENNHAEBC	DTCRSNACTV	YCHADSAYLC	MSCDAQVHSA	NRVASRRHKRV	60
Col1	MLKVES...	...NWAQAC	DTCRSAACTV	YCRADSAYLC	SSCDAQVHAA	NRLASRRHERV	52
Col2	MLKEESNESG	...TWARAC	DTCRSAACTV	YCEADSAYLC	TTCDARVHAA	NRVASRRHERV	56
Consensus	MLK - ESN - - G	- - - NWARAC	DTCRSAACTV	YC - ADSAYLC	- SCDAQVHAA	NRVASRRHERV	60
Constans	RVCESCCERAP	AAFICEADDA	SLCTACDSEV	HSANPLARRH	QRVPILPISG	NSFSSMTTTH	120
Col1	RVCQSCCERAP	AAFECKADAA	SLCTTCDSFI	HSANPLARRH	QRVPILPISG	YSYSS-TATN	111
Col2	RVCQSCCERAP	AAFLCKADAA	SLCTACDAEI	HSANPLARRH	QRVPILPISG	NSCESMAFS	115
Consensus	RVCQSCCERAP	AAFLCKADAA	SLCTACDSEI	HSANPLARRH	QRVPILPIS -	NS - SSMT - T -	120
Constans	HQSEKTMTD	EKRLLVVDQEE	GEEDKDAKE	VASWLLPNSD	KNNNNNONNGE	LFSDDEYLNLV	180
Col1	HSCETTVTD	ENRLLVLGQEE	EDDEDAEA	ASWLLPNSG	KMSGN - NNGPF	SI GDEYLNLV	167
Col2ET	DADNDEDDRE	VASWLLPNPG	KNIQNQNGPF	LFGEVEYLNLV	157
Consensus	H - - E - T - TDP	E - RLV - - QEE	- - E - D - DA - E	VASWLLPNSG	KN - GNQNGPF	LFGEVEYLNLV	180
Constans	DYNSSMDYKFF	TGEYSQHQQN	CVPQTSYGG	DRVVPLKLEE	SRGHQCHNQQ	NFQFN I KYG	239
Col1	DYSSSDKQFF	TDQSNQYQLD	CNVQRSYGE	DGVVPLQIEV	SKG - MYQEQQ	NFQLSINCG	224
Col2	DYSSSMDNQFF	EDN1 - - - Q	VTYQRSFEGG	DGVVPLQVEE	STSHLQSSQQ	NFQLGINYGF	211
Consensus	DYSSSMD - QFF	TD - - - Q - Q -	C - VPQRSYGG	DGVVPLQ - EEE	S - GH - - Q - QQ	NFQL - INYGF -	240
Constans	SSGTHYNDNG	SLNHNAYE	SSMELTGVVPE	STACVTTASH	PRTPKGTVEQ	QPDPA SQMIT	297
Col1	SSGALRSSNG	SLSHMNVNV	SSMDLGVVPE	STTSDATVSN	PKSPKAVTDO	PPYPFAQM	280
Col2	SSGAHYNNNS	LKDLNHSASV	SSMDISVVPE	STASDI TVQH	PRTTKETIDQ	LSGPPPTQV	269
Consensus	SSGAHYN - NG	- - SLNH - A - V	SSMD - GVVPE	STASD - TVSH	PRTPR - T - DQ	- P - PP - QM -	300
Constans	VTQLSPMDRE	ARVLRVYREKK	KTRKFEKTIR	VASRKAYAEI	RPRVMGRFAK	REIEAEE	355
Col1	...LSPRDRRE	ARVLRVYREKK	KMRKFEKTIR	VASRKAYAEK	RPRIKGRFAK	KKDVDEEANQ	337
Col2	VQLTPMERE	ARVLRVYREKK	KTRKFEKTIR	VASRKAYAEI	RPRIKGRFAK	RIETEAEAE	329
Consensus	V - QLSPMDRE	ARVLRVYREKK	KTRKFEKTIR	VASRKAYAEI	RPRIKGRFAK	R - - - EEA - Q	360
Constans	GFNTMLMYNT	GYGIVPSPF373					
Col1	AFSTMITFD	GYGIVPSPF355					
Col2	IFSTSLMSET	GYGIVPSPF347					
Consensus	- FSTMLM - T	GYGIVPSPF378					

Figure 3.8 Prettybox comparison of the predicted CONSTANS protein with CONSTANS-LIKE1 (Col1, Putterill *et al.*, 1997) and CONSTANS-LIKE2 (Col2, Ledger *et al.*, 1996).

using TBLASTX against all sequences released during July 1997. It is located on the non-annotated BAC, AC002332, which is located near marker ve015 at map position 63.3 on chromosome 2. It is also homologous to *CO* over the zinc fingers, with 37% identity. Recently a BLAST search was performed using only the basic carboxyl domain of the *CO* protein. Regions homologous to this domain were found near to, and in the same orientation as, both of the sequenced genes described above. These domains are assumed to be part of the same two *CO* homologues and have been included in the analysis of homology. The third homologue identified, AC000104, was found on BAC F19P19 at the top of chromosome 1 using the basic, C-terminal domain in the BLAST search. The homology detected by the BLASTP comparison extends over a 22 amino acid region, containing the residues affected in alleles *co-5* and *co-7*, with 14 identical and a further 2 similar amino acids (73%). The homologue also contains a putative nuclear localization signal in a position similar to that found in *CO* but this was not detected by the BLASTP comparison. More recent searches have revealed homology to the *CO* zinc fingers at the N-terminus of the predicted amino acid sequence of this gene. However, this has not been included in the homologue sequence comparison as the region of homology is small and the annotation of this BAC was incomplete at the time of writing. AC000104 is interesting because it is the only homologue which appears to contain both the putative nuclear localization domain and the more highly conserved domain separated by a spacer of 13 amino acids. In *CO* these two domains are directly adjacent to each other. This is further evidence that the C-terminal basic region can be subdivided into two distinct domains which may have different functions.

The BLASTP search of all available databases found some less well conserved proteins from other species which have not been included in the main sequence comparison but have been compared independently to *CO* (Figure 3.10) The conservation between *CO* and these proteins seems to be confined to the conserved cysteines and histidines in the zinc fingers,

in addition to a small number of the most highly conserved hydrophobic residues also in this domain. It is possible that these proteins could have a role in binding zinc through these domains but are only very distantly related to CO. The proteins included the GTP-binding protein ARD1 from humans (Mishima *et al.*, 1993; Vitale *et al.*, 1998) and the equivalent ARD1-like protein from Rat, and the putative transcription factor CeLIM-4 from *Caenorhabditis elegans* (Accession U72348, Hobert *et al.*, unpublished results).

Using the TBLASTX search against the GenBank Expressed Sequence Tag (EST) database using either the whole *CO* genomic sequence, or the basic C-terminal domain, a total of 56 ESTs with homology to *CO* in the correct frame were detected. One EST (T76837) did not show homology to either the zinc finger region or the basic C-terminal domain but to the middle of the protein. Comparison using GAP (GCG) to the known homologues suggested that T76837 was a partial clone derived from the *COL1* gene. Comparison of the predicted protein sequence of the remaining 55 ESTs to each other and to the known *CO* homologues allowed them to be classified into 27 groups, 16 with homology to the zinc fingers and 11 with homology to the basic domain. Each group is thought to represent a unique gene, although the poor quality of some EST sequences sometimes made their classification difficult. The groups are shown in Tables 3.2 and 3.3. Two ESTs (N38572 and H36917) derived from the previously described *CO* homologue, *STO* (Lippuner *et al.*, 1996) were found. These two ESTs were also detected when *STO* itself was used in a database search (Lippuner *et al.*, 1996). No *CO* ESTs were detected, but this is not unexpected as *CO* is expressed at a very low level.

At the N-terminus of the protein in the zinc finger region the Arabidopsis ESTs probably represent nine genes. Seven of these were compared over the zinc fingers using the Pileup and Prettybox programs (GCG) to CO and the five known genes homologous in this region (Figure 3.9; Table 3.2). The two remaining ESTs (N65760 and R30042) which are also possibly unique were not included because of very poor sequence data. The rice ESTs

Zinc Finger 1

	*																				*																												
Constans	C	D	T	C	R	S	N	A	C	T	V	Y	C	H	A	D	S	F	Y	L	C	M	T	S	C	D	A	Q	V	H	S	A	N	R	V	A	S	R	H	K	R	V	R	V					
Col2	C	D	T	C	R	S	S	A	A	C	T	V	Y	C	H	A	D	S	F	Y	L	C	C	T	S	C	D	A	Q	V	H	S	A	A	N	R	V	A	S	R	H	K	R	V	R	V			
Col11	C	C	D	T	C	R	S	S	A	A	C	T	V	Y	C	H	A	D	S	F	Y	L	C	C	C	S	C	D	A	Q	V	H	S	A	A	N	R	V	A	S	R	H	K	R	V	R	V		
Sto	C	C	D	T	C	R	S	S	A	A	C	T	V	Y	C	H	A	D	S	F	Y	L	C	C	C	S	C	D	A	Q	V	H	S	A	A	N	R	V	A	S	R	H	K	R	V	R	V		
Z97338	C	C	D	F	C	G	T	E	K	A	L	L	Y	C	K	S	D	S	A	K	L	L	C	C	C	L	P	C	D	Q	H	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
Ac002332	C	C	E	F	C	G	T	E	T	A	V	L	F	C	R	A	D	T	A	K	L	L	C	C	C	L	P	C	D	Q	H	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
N96038	C	C	D	F	C	G	T	E	K	A	L	L	Y	C	K	S	D	S	A	K	L	L	C	C	C	L	P	C	D	Q	H	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
▷N96950	C	C	N	x	C	K	A	E	A	T	L	V	F	C	C	A	D	E	A	A	L	L	C	C	C	G	A	C	x	E	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
T04394	C	C	K	F	C	K	A	Y	R	A	V	V	Y	C	C	I	A	D	T	A	N	L	C	C	C	W	A	C	x	E	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
▷T04667	C	C	x	S	C	V	K	R	R	A	R	W	Y	C	C	I	A	D	T	A	N	L	C	C	C	H	A	C	x	E	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
T13953	C	C	D	S	C	K	S	A	T	A	A	L	Y	C	C	R	P	X	A	A	F	L	C	C	C	L	A	C	D	S	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
T44486	C	C	x	A	C	K	S	V	T	A	A	L	V	F	C	C	R	P	X	A	A	F	L	C	C	C	L	A	C	D	S	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V
◇T75904	C	C	D	Y	C	D	S	S	V	A	L	V	Y	C	C	K	A	D	S	A	K	L	C	C	C	L	A	C	D	S	K	L	H	T	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
c19632	V	Y	C	C	R	A	D	A	A	R	L	C	C	X	Q	C	D	N	R	X	V	H	G	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
◇C23555	R	A	A	V	V	Y	C	C	R	A	D	A	A	R	L	C	C	X	Q	C	D	N	R	X	V	H	G	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
D21975	R	A	A	V	V	Y	C	C	R	A	D	A	A	R	L	C	C	X	Q	C	D	N	R	X	V	H	G	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
D24180	C	D	V	C	A	A	E	P	A	A	V	L	C	C	A	D	E	A	A	L	L	C	C	C	S	A	C	C	D	N	R	R	V	H	R	A	A	N	K	L	A	S	R	H	K	R	V	R	V
◇D24247	C	D	A	C	x	S	A	A	A	A	V	V	V	C	C	A	D	E	A	A	L	L	C	C	C	A	A	C	x	V	E	M	H	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
D25007	C	N	A	C	G	A	A	E	A	R	V	L	C	C	A	D	E	A	A	L	L	C	C	C	T	A	C	C	D	E	E	V	H	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
▷D25039	C	x	F	C	R	E	Q	R	S	M	V	Y	C	R	S	D	A	A	S	L	L	C	C	C	L	S	C	D	E	R	N	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V	
▷D41222	C	D	S	C	R	S	A	P	C	G	x	Y	x	L	A	D	S	G	T	L	L	C	C	C	A	T	C	D	A	D	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
▷D47287	C	D	S	C	R	S	A	P	C	G	x	Y	x	L	A	D	S	G	T	L	L	C	C	C	A	T	C	D	A	D	V	H	S	A	A	N	K	L	A	S	R	H	K	R	V	R	V		
Consensus	C	-	-	C	-	-	-	-	-	A	-	V	-	C	-	A	D	-	A	-	L	-	C	-	-	C	D	-	-	V	H	-	A	N	-	L	A	-	R	H	-	R	-	-	-	-			

Zinc Finger 2

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Figure 3.9 Prettybox comparison of the predicted protein sequence of the *CO* zinc fingers with a number of homologues in the GenBank database.

For a summary of the nature and origin of the homologues see Table 3.2 overleaf.

A spacer of between 9 and 18 amino acids between the two zinc fingers in six of the homologues (STO and five of the Rice ESTs) is not included in the lineup.

- ◇ These homologues had very incomplete sequence for zinc finger 2 but this is not included in the lineup.
- ▷ These homologues had sequence information for only one zinc finger so these single fingers have been lined up with both zinc finger 1 and 2 of CO.

* The amino acids affected by the mutations in the *co* alleles.

Origin	Representative gene/EST	ESTs probably derived from the same gene	Species
Cloned by chromosome walking	<i>CONSTANS</i>	None	<i>Arabidopsis thaliana</i>
Maps 3 kb from <i>CO</i>	<i>CONSTANS-LIKE1</i> (Putterill <i>et al.</i> , 1997)	T76837	<i>Arabidopsis thaliana</i>
Genomic library	<i>CONSTANS-LIKE2</i> (Ledger <i>et al.</i> , 1996)	None	<i>Arabidopsis thaliana</i>
cDNA conferred salt tolerance to yeast	<i>SALT TOLERANCE (STO)</i> (Lippuner <i>et al.</i> , 1996)	N38572, H36917	<i>Arabidopsis thaliana</i>
Sequenced and annotated BAC	Z97338	None	<i>Arabidopsis thaliana</i>
	AC002332	None	<i>Arabidopsis thaliana</i>
EST	N96038	H76905	<i>Arabidopsis thaliana</i>
	N96950	Z34081	<i>Arabidopsis thaliana</i>
	T04394	T04563, T46626	<i>Arabidopsis thaliana</i>
	T04667	None	<i>Arabidopsis thaliana</i>
	T13953	None	<i>Arabidopsis thaliana</i>
	T44486	H37351, H7329, H37479, T44465, T45830, T21196, N65050, T2131, T44868, T45557, N37164	<i>Arabidopsis thaliana</i>
	T75904	None	<i>Arabidopsis thaliana</i>
	C19632	None	Rice
	C23555	C19632, D47852	Rice
	D21957	None	Rice
	D24180	None	Rice
	D24247	None	Rice
	D25007	None	Rice
	D25039	None	Rice
	D41222	None	Rice
	D47287	D25007	Rice

Table 3.2 The *CO* homologues with similarity over the zinc fingers. Each group, represented by a single member in the comparison in Figure 3.9, is thought to represent a unique gene. The ESTs and genomic homologues identified during the *Arabidopsis* sequencing programme are identified by their GenBank accession numbers.

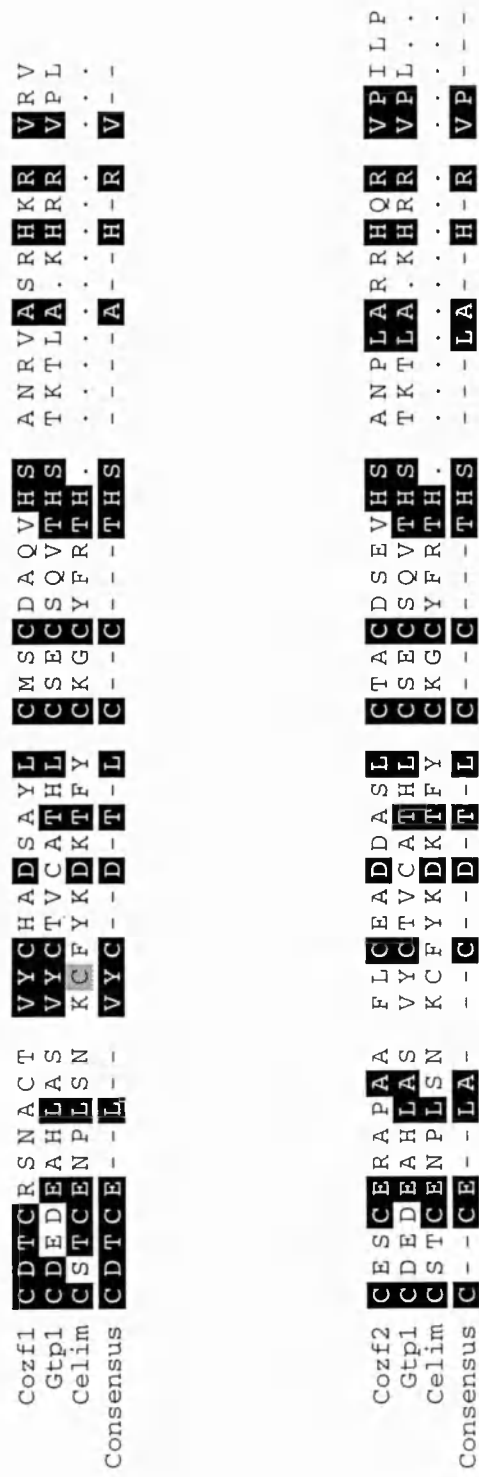


Figure 3.10 Prettybox comparison of the predicted protein sequence of the CONSTANS zinc fingers with two proteins in the Swissprot database. CONSTANS zinc finger 1 (Cozf1) and zinc finger 2 (Cozf2) are compared independently with the two proteins. Gtp1 *Homo sapiens* GTP-binding protein ARD1 (Mishima *et al.*, 1993). Celim *Caenorhabditis elegans* putative transcription factor CelIM-4 (Hobert *et al.*, unpublished results, Accession U72348)

detected also probably represent nine unique genes and these are included in the sequence comparison. The predicted protein sequences of all of the ESTs were edited to remove poor sequence particularly at either end of the sequenced region. Where only one zinc finger is present in a homologue it is unclear whether this represents a subgroup of CO homologues that contain a single finger or whether this is due to insufficient sequence data. In some cases it was difficult to establish whether the partial sequence was derived from zinc finger 1 or zinc finger 2, because of the high degree of similarity between the CO fingers. These ESTs have been compared to each zinc finger independently.

The high number of related genes indicates that *CO* is one representative of a substantial gene family in *Arabidopsis*. The 14 new members identified by the methods described here represent only a part of the family which might contain many more genes that have not yet been identified. ESTs were not detected for *CO* or 3 of the 5 known genes homologous over the zinc fingers. The identification of 9 *CO* homologues that probably represent different genes in rice indicates that this gene family is not restricted to *Arabidopsis*.

Eight unique *Arabidopsis* ESTs were identified with homology to the basic domain at the C-terminus. These were compared to *CO* and the 5 known genes which are homologous in this domain using Pileup and Prettybox as before (Figure 3.11). In addition a rice EST (D46441), a maize EST (T12686) and a pine EST (H75099) were also identified. All homologues have the very highly conserved basic domain, defined in *CO* by the region between isoleucine 326 and lysine 347, but the preceding NLS is missing in three out of the 11 ESTs, including the pine homologue. The lack of reliable and extensive sequence data may account for this observation, but it is possible that like homologue AC000104, this domain may exist in some genes in a different part of the protein. The two other homologues identified from the *Arabidopsis* genomic sequencing programme, Z97338 and AC002332, probably have potential NLSs in regions upstream of their very highly conserved basic domains. However, in the case of AC002332 it was not detected in any

Origin	Representative gene/EST	ESTs probably derived from the same gene	Species
Cloned by chromosome walking	<i>CONSTANS</i>	None	<i>Arabidopsis thaliana</i>
Maps 3 kb from <i>CO</i>	<i>CONSTANS-LIKE1</i> (Putterill <i>et al.</i> , 1997)	None	<i>Arabidopsis thaliana</i>
Genomic library	<i>CONSTANS-LIKE2</i> (Ledger <i>et al.</i> , 1996)	None	<i>Arabidopsis thaliana</i>
Sequenced and annotated BAC	Z97338	None	<i>Arabidopsis thaliana</i>
	AC000104	None	<i>Arabidopsis thaliana</i>
	AC002332	None	<i>Arabidopsis thaliana</i>
EST	F20076	None	<i>Arabidopsis thaliana</i>
	H76545	H76952	<i>Arabidopsis thaliana</i>
	N38693	None	<i>Arabidopsis thaliana</i>
	N65003	None	<i>Arabidopsis thaliana</i>
	T76791	AA394422	<i>Arabidopsis thaliana</i>
	W43720	Z34786	<i>Arabidopsis thaliana</i>
	Z17946	T20548	<i>Arabidopsis thaliana</i>
	Z37717	AA042520, Z24476	<i>Arabidopsis thaliana</i>
	D46441	None	Rice
	T12686	None	Maize
	H75009	None	Pine

Table 3.3 The *CO* homologues with similarity over the basic, C-terminal domain. Each group, represented by a single member in the comparison in figure 3.11, is thought to represent a unique gene. The ESTs and genomic homologues identified during the *Arabidopsis* sequencing programme are identified by their GenBank accession numbers.

BLAST search and does not appear to be in a region resembling an ORF. Neither have been included in the analysis at this point because of the incomplete annotation of the BACs at the time of writing.

In addition to the two subdomains of this basic region there is some conservation between homologues at the very C-terminus of the protein. CO, COL1 and COL2 have a short acidic stretch to the carboxyl side of the basic domain and a completely conserved hydrophobic motif of 8 amino acids at the end of the protein (GYGIVPSF). There is partial conservation of the hydrophobic motif in EST N38693 (GFGVVPSF), although again there is insufficient reliable sequence in this region for many of the ESTs.

A summary of the important residues and structural domains of the CO protein, identified from the molecular analysis of the mutations in the seven alleles and the conservation between homologues, is shown in Figure 3.12.

3.4 Discussion

The seven *constans* alleles were characterized at the molecular level and the mutations identified. All are predicted to be missense mutations which affect a single amino acid except for the X-ray-induced allele *co-1*, where a deletion precisely removes three amino acids leaving the remainder of the protein unaffected. The mutations were found to be clustered in two highly conserved parts of the CO protein which are assumed to define important functional domains.

As described in the introduction, the CONSTANS protein shares some structural similarities but little sequence homology with the DNA binding domain of the GATA-1 family of transcription factors. Near the amino terminus of the CO protein are two potential zinc finger modules of the C₂C₂ type that are separated by 19 amino acids. Comparison of the

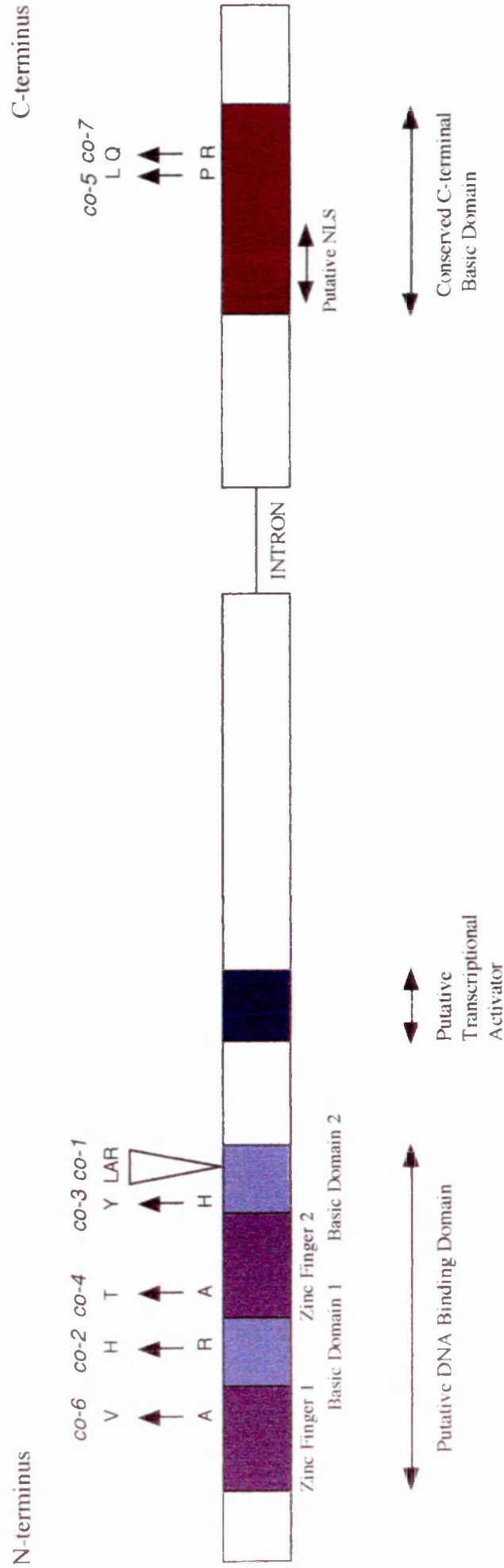


Figure 3.12 A summary of some of the structural features of the CO protein thought to be important for its function. The predicted amino acid substitutions in six of the mutant alleles and the predicted deletion of three amino acids in *co-1* are shown.

two potential fingers, and the regions to the carboxyl side of each, show significant homology, with 46% identity and 86% similarity (Putterill *et al.*, 1995). The most conserved regions are found at the carboxyl side of each finger. These regions contain a number of basic amino acids which, as well as the conserved spacing of cysteines, is another structural feature of the DNA binding domain common to the GATA-1 transcription factors (Romain *et al.*, 1993).

Five of the seven alleles contain mutations which affect the potential DNA binding domain. Three of the mutations affect the basic domains on the carboxyl side of each of the two zinc fingers; *co-2* adjacent to zinc finger 1, *co-1* and *co-3* adjacent to zinc finger 2.

The three amino acids deleted from *co-1* are highly conserved. In five out of eleven unique homologues the motif is completely conserved and differs by one or two amino acids in the other six. The first leucine deleted from *co-1* is conserved in all eleven homologues which have this domain. Many CO homologues also have this, or a very similar motif, in the equivalent position adjacent to zinc finger 1.

The histidine residue which is affected in the *co-3* mutant is extremely conserved in related genes, being present in 14 out of 14 homologues. At an equivalent position at the carboxyl side of zinc finger 2 there is also a histidine conserved in all 22 homologues. The position of histidines 48 and 91 in relation to the conserved cysteine pairs, the conservation of these histidines among even distantly related homologues and the severity of the *co-3* mutant may suggest an even more fundamental role in the zinc binding structure. Alternative structures for the zinc binding domain are discussed later.

The three dimensional solution structure of the DNA binding domain of the chicken erythroid transcription factor, cGATA-1, was determined using multidimensional heteronuclear magnetic resonance spectroscopy (Omichinski *et al.*, 1993). In cGATA-1 a 59 amino acid region comprising the carboxyl zinc finger and its basic tail has been defined

as the minimal DNA binding unit. The analysis revealed an alpha helix involved in interaction with DNA in the major groove, encompassing the second cysteine pair and seven additional amino acids to the carboxyl side of the zinc finger. A smaller alpha helix which interacts with DNA in the minor groove lies at the beginning of the long basic tail. The predicted alpha helix in CO, affected by the deletion in *co-1*, was found in the conserved region on the carboxyl side of zinc finger 2, and might suggest a role for this region in interaction with DNA.

In addition to its role in DNA binding, this region of GATA-1 is also implicated in protein-protein interactions, mediating the self-association of the protein or interaction with other members of the GATA family (Crossley *et al.*, 1995). If the equivalent region in CO also has multiple roles, the mutations in *co-1* and *co-3* may each affect a different protein function.

Figure 3.13 shows a model for the structure of the zinc finger region of CO, based on the C₂C₂ type zinc fingers of the GATA transcription factors. The structure of the zinc finger region of cGATA-1 and the functions of subdomains and individual residues within the zinc fingers are included for comparison.

The mutation in *co-2* affects basic domain 1, between the two zinc fingers. The arginine residue which is substituted by a histidine in *co-2* is conserved in all 20 homologues for which sequence information at this position is available. The arginine residue lies close to zinc finger 2, four amino acids from the first cysteine pair. In cGATA-1 this region is important for maintaining the structural integrity of the zinc binding region (Omichinski *et al.*, 1993). There is also a conserved arginine in an equivalent position in basic domain 2. Unlike cGATA-1, (Yang and Evans, 1992) the two basic domains to the carboxyl side of each zinc finger share a considerable amount of homology (Putterill *et al.* 1995). This may implicate the zinc finger 1 basic domain, affected in *co-2*, in interaction with DNA in

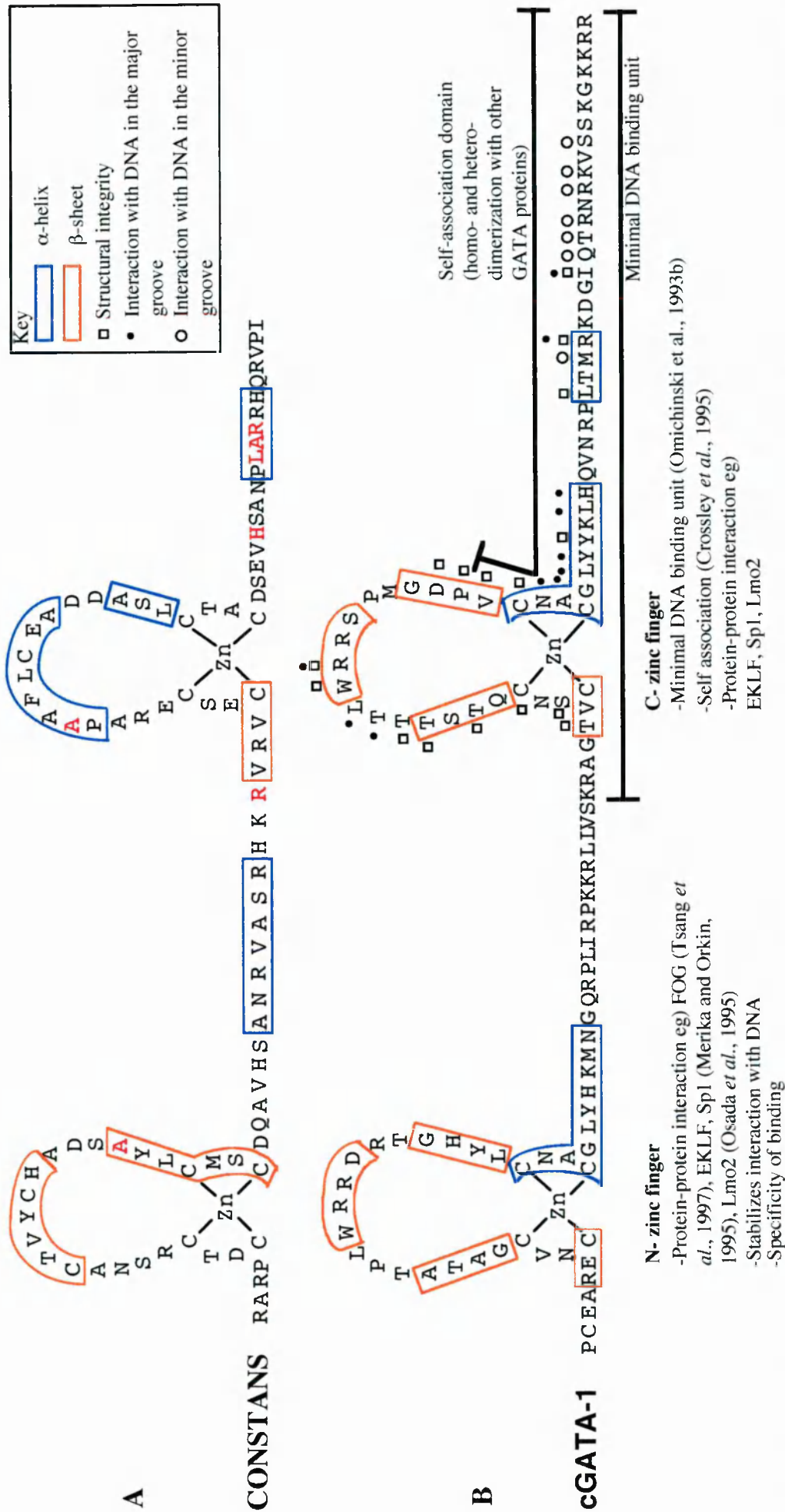


Figure 3.13 A) Sequence of the CONSTANS zinc finger region with secondary structure predicted by the PHD predict protein program (Rost, 1996). The region has been represented as two zinc fingers of the CX₂C type, based on the DNA binding domain of the GATA-1 transcription factors. The amino acids affected by the mutations in the *co* alleles are in red. B) Sequence of cGATA-1 (chicken). The secondary structure, the identity of the zinc-binding residues and the role of other residues in stability or interaction with DNA was resolved by NMR spectroscopy (Omichinski *et al.*, 1993a).

a similar manner to that proposed for the zinc finger 2 basic domain. Alternatively, this region may be involved in transactivation. At the equivalent region in the transcription factor c-GATA-3, substituting alanine residues for three basic residues (KRR) resulted in a protein that could not activate transcription, although DNA binding was not affected (Smith *et al.*, 1995).

The remaining two *co* alleles affect residues in the loop of each of the two zinc fingers. The weakest mutant, *co-4*, has an amino acid substitution in zinc finger 2. The alanine residue affected is the least conserved of all amino acids affected in the alleles, being present in 12 out of 19 homologues. A mutation at this position seems to be more tolerated than mutations at positions affected in the other alleles, suggesting that this residue is less important for structure or function of this domain. However, the nature of the mutation is also likely to be important; a change to the weakly related threonine (Taylor, 1986) may not disrupt the function of the protein as radically as a change to a very different amino acid. This is further demonstrated by the presence of a threonine at this position in the CO homologue Z97338.

The mutation in *co-6* also affects an alanine residue but in zinc finger 1. This alanine residue, however, is highly conserved, being present in 18 out of 19 homologues. The mutation affects two predicted β sheets (β -hairpin) which are also a structural feature of the finger region of the DNA binding domain of cGATA-1. It is possible, therefore, that the region affected in *co-6* is important for maintaining the structure of the N-terminal zinc finger.

If CO is related functionally to the GATA-1 transcription factors it is likely that the two fingers have distinct roles. The analysis of a series of mutations generated in the zinc finger domain of c-GATA-1 demonstrated that the C-terminal zinc finger and adjacent basic tail was sufficient to bind specifically to the GATA motif in DNA. The N-terminal zinc finger

was found to play a role in DNA binding by stabilizing the DNA-protein complex (Yang and Evans, 1992). Recently, a role for the N-terminal finger of GATA transcription factors in protein-protein interaction has been proposed (Reviewed in Mackay and Crossley, 1998). A two-hybrid screen with GATA-1 identified FRIEND OF GATA (FOG), a Krüppel-like protein with nine zinc fingers. Finger 6 of FOG specifically interacts with the N-terminal finger of GATA-1 (Tsang *et al.*, 1997). A number of other proteins are known to interact with the finger region of GATA proteins, for example, the Krüppel-like transcription factors Sp1 and EKLF, which can interact with either finger 1 or finger 2 (Merika and Orkin, 1995) and the LIM-domain protein Lmo2/RBTN2 (Osada *et al.*, 1995) which forms a complex with GATA-1.

Experiments are currently underway to ascertain whether the wild type CO protein is involved in binding DNA and to identify target sequences. A two-hybrid screen to identify proteins that might interact with the CO zinc fingers is also in progress. Similar analysis of the mutant proteins may help towards understanding the role of the zinc finger domain and the effect of the mutations upon its structure and function.

Several alternative structures for the potential zinc binding motif in the sunflower gene SF3 were originally proposed in view of the presence of a number of potential zinc binding ligands (Baltz *et al.*, 1992). The conservation of the spacing of some of these residues classified this protein as a member of the LIM family of cysteine-rich zinc binding proteins (Sanchez-Garcia and Rabbitts, 1994). This suggested which potential ligands might be involved in zinc binding; C-X₂-C-X₁₇-H-X₂-C-X_{2/3}-**C/G**-X₁₇-C-X₂-H, with a cysteine or glycine residue (in bold) present in the third pair of ligands in finger 1 and finger 2, respectively.

Analysis of the CONSTANS protein sequence, in particular the conservation of residues in the putative zinc binding region, suggest alternate configurations particularly in light of

the extra conserved cysteine and histidine residues close to the conserved cysteine pairs. As already discussed, the cysteine pairs have a similar spacing to those found in the GATA-1 family but in CO there are extra well conserved potential zinc ligands. Each finger has the spacing C-X₂-C-X₈-C-X₇-C-X₂-C-X₄-H-X₈-H and all these residues are conserved in all closely related CO homologues found so far. The most severe mutant, *co-3*, has a point mutation which affects the highly conserved histidine in zinc finger 2. It is possible that the histidines in each zinc finger could form a mixed second pair of ligands with a cysteine, and the conserved unpaired cysteine between the cysteine pairs could also be involved in zinc binding which would suggest a cluster arrangement rather than a classical C₂C₂ finger. However, this arrangement of conserved cysteines and histidines is not reminiscent of any known zinc binding motif. Using a site-directed mutagenesis approach it may be possible to target specific residues in the zinc fingers and ascertain which of these are ligands directly involved in zinc binding.

In addition to the putative DNA binding domain there is another highly conserved region at the C-terminus of the CO protein. The domain starts with a leucine residue at position 301 and runs to the end of the protein (phenylalanine 373). Within this region there are small islands of higher conservation. This domain does not have strong similarity to those of previously described proteins, except for the CO homologues, but it has a number of features which could suggest functions. Mutant alleles *co-5* and *co-7* affect residues within this region.

The weak *co-5* mutant phenotype is caused by the substitution of a leucine for a proline. This residue is less well conserved in CO homologues than the residues affected in the stronger alleles. The proline residue affected is conserved in 10 out of 15 homologues. At this position, in four of the remaining homologues a basic amino acid (arginine or lysine) is substituted for proline indicating the importance of the presence of positively charged

residues in this domain. The homologue Z97338, like *co-5*, contains a leucine at this position. This is similar to the case of allele *co-4* (alanine to threonine substitution) where the same homologue contained a threonine at the position affected by *co-4*. This suggests that the substitutions in these two weak alleles can be tolerated at this position in closely related genes, and therefore presumably do not greatly affect the function of these conserved domains.

The substitution of glutamine for arginine 340 causes a severe late-flowering phenotype in the *co-7* mutant. This residue is highly conserved, being present in 16 out of 16 homologues. The net positive charge of this region, and its strong conservation, suggests that a reduction in this charge, for instance by the introduction of the non-charged glutamine, may have a deleterious effect on the structure or function of the domain.

The most striking feature of this C-terminal domain is the preponderance of basic amino acids, giving the entire region a net charge of +9. The most conserved stretch of 44 amino acids from aspartic acid 305 to arginine 348 has a net charge of +13. When the CO amino acid sequence was compared to the secondary structure prediction program "PHD predict protein" (Rost 1996) the basic domain was predicted to form two alpha helices from residues 306 (arginine) to 341 (valine) and residues 344 (Arginine) to 354 (Glutamic acid). The amino acid substitutions in both *co-5* and *co-7* are predicted to reduce the number of residues in the short loop or turn between the two alpha helices. If the orientation of particular residues in this structure is important, for example in interacting with another molecule, the disruption to the angle of a helix may be sufficient to destabilize the interaction.

The association of both charged clusters and regions able to form alpha helices with functional domains of transcription factors is well documented (Brendel and Karlin 1989, reviewed in Ptashne, 1988 and for plant transcription factors Meshi and Iwabuchi, 1995).

Basic domains are often associated with DNA binding domains, where the positively charged amino acid side-chains interact with the negatively charged phosphate backbone of DNA. In the basic-region leucine-zipper (bZIP) family and the basic-region helix-loop-helix (bHLH) family the basic domain forms an alpha helix which interacts with DNA in the major groove (Glover and Harrison, 1995; Ma *et al.*, 1994). A basic domain is also involved in the interaction with DNA in the MADS-box family (Bruhn *et al.*, 1994). The association with the zinc fingers in the GATA family of transcription factors of a basic domain involved in the interaction with DNA in both the major and minor grooves has already been discussed.

A more transient protein-DNA interaction is proposed for a basic, highly conserved domain found in the maize VP1 transcription factor (Hill *et al.*, 1996). A 40 amino acid peptide containing the basic domain BR2 from VP1 is able to enhance the binding of the wheat bZIP transcription factor EmBP-1 to the *Em* promoter *in vitro*. The BR2 domain has the potential to form an alpha helix, with possible DNA binding properties. However, a stable VP1-DNA complex is not detected in gel-shift experiments, and it is proposed that the interaction of the basic domain of VP1 is transient, possibly inducing a change in DNA conformation, and allowing other transcription factors to bind to their target sites. If this region of CO is involved in DNA binding then together with the zinc fingers CO might be predicted to contain two DNA binding regions. Transcription factors with more than one putative DNA binding domain do exist; for example, the twin autonomous domains of the rice protein GT-2 (Dehesh *et al.*, 1992); the presence of both a homeodomain and zinc finger-motifs in a *Drosophila* protein, (Fortini *et al.*, 1991). However, whether the CO protein is involved at all in DNA binding, through the zinc fingers or any other domain, has yet to be determined.

In certain transcriptional repressors basic motifs are thought to act as the repressor domain; for example, the *E. coli*-encoded sequence SSB24, which can act as a potent

repressor of transcription in yeast when fused to the DNA binding domain of GAL4 (Saha *et al.*, 1993). However, as CO can activate transcription in the same yeast system (M. Costa, unpublished results) it seems unlikely that it has a role in repression.

Alternatively, the C-terminal domain of CO might act to activate transcription. The formation of an alpha helix is thought to be important in acidic activation domains (Franken *et al.*, 1994) where the solvent-exposed residues are predominantly negatively charged and are thought to provide a surface for the interaction with another protein in the transcriptional machinery. Although this region of CO is positively charged, a genetic screen in yeast has identified potential transcriptional activating sequences from tobacco, which are characterized by a net positive charge (Estruch *et al.*, 1994). One of these was found to be homologous to the BBC-1 human protein, a member of a conserved family of basic hydrophilic proteins which contain potential NLSs in a region with a preponderance of basic amino acids (Sáez-Vásquez *et al.*, 1993).

The prediction that this domain may form an alpha helix could implicate it in the interaction with another protein not necessarily directly involved with transcriptional activation. It is possible, for example, that CO could bind DNA as a hetero- or homodimer, and this domain could mediate or stabilize the interaction with its partner.

One possible role for at least part of this basic domain could be nuclear localization. If CONSTANS is a DNA binding protein it must be localized to the nucleus. Most nuclear proteins carry a nuclear localization signal (NLS) which directs the protein through the nuclear pore into the nucleus (Garcia-Bustos *et al.*, 1991; Raikel, 1992). A Prosite search with the CONSTANS protein sequence did not detect any potential NLS motifs (Putterill *et al.*, 1995). Recently, however, *CONSTANS* and *CONSTANS-LIKE 1* orthologues from *Brassica napus* have been isolated and sequenced (Robert *et al.*, 1998; Chapter 6). The *Brassica napus* orthologue of *CONSTANS-LIKE 1* (*COL1* Putterill *et al.*, 1997) has three conservative amino acid changes in a basic motif compared to CO, and this is detected by

Prosites as a potential bipartite NLS (L. Robert, unpublished results). A bipartite NLS consists of two basic amino acids, a spacer region with more than four amino acids (but typically ten) and a basic cluster in which three out of five residues must be basic. This type of NLS was first defined in the *Xenopus* nuclear protein, nucleoplasmin (**K**RPAATKKAGQA**KKKK**, with the residues important for nuclear targeting in bold; Robbins *et al.*, 1991). The potential *COL1 B. napus* orthologue NLS is **R**KARVMRYRE**KKK**TRK. In CONSTANS, the equivalent sequence reads **R**EARVLRyRE**KRK**TRK. In the first domain, only one residue is basic, and this divergence from the consensus, along with the two other differences probably prevented the Prosites program from detecting the NLS in CO. However, this region may still represent a functional NLS. Preliminary results indicate that the CO protein is localized to the nucleus (M. Costa, unpublished results).

The role of the remainder of this domain, including the region affected in the alleles, could be further investigated by fusing serial deletions of the CO protein to the DNA binding domain of the GAL4 transcription factor, in a yeast two-hybrid system. This may reveal proteins which interact with CO, possibly through the conserved basic domain.

If CONSTANS plays a role in the regulation of transcription it might contain a transcriptional activation or repression domain. Three main activation domains characterized by different sequence motifs have been reported (reviewed by Mitchell and Tjian, 1989) and although they do not appear to have a strong consensus sequence they have recognisable characteristics. The CONSTANS protein contains five regions with similarity to transcriptional activation domains; two acidic domains and three glutamine-rich domains. However, the latter are between 9 and 14 residues long and therefore are much shorter than published examples such as the 143 amino acid glutamine-rich domain B of the mammalian transcription factor Sp1 (Courey and Tjian, 1988). The CO protein can activate transcription of reporter genes in yeast, when fused to the DNA binding domain of GAL4 (M. Costa,

unpublished results). It is more common for activation by a foreign transcription factor in a yeast system to occur through an acidic activation domain rather than a glutamine-rich or proline-rich activator (Schwechheimer *et al.*, 1998). This may suggest that CO contains an acidic domain that can activate transcription. The most convincing candidate is the acidic domain which extends approximately from glutamic acid 124 to glutamic acid 150. Experiments are currently underway to establish whether this acidic region, or other domains are responsible for activation.

Zinc finger transcription factors often exist as families of related genes. In the human genome, the TFIIII family of zinc binding transcription factors, characterized by the classic C₂H₂ zinc fingers, reportedly contains up to 1000 members (Bellefroid *et al.*, 1990). Recently, nine members of a diverse family of *Arabidopsis* C₂H₂ zinc finger proteins were isolated using degenerate PCR primers designed to highly conserved regions in the petunia *EPF1* zinc finger gene (Meissner and Michael, 1997). A similar approach was used to isolate eight *Arabidopsis* single fingered C₂H₂ genes, thought to represent the majority of a subfamily containing up to 10 members (Tague and Goodman, 1995). Since the *CO* gene was first identified, evidence of an extensive family of related zinc finger proteins in *Arabidopsis* has been growing (Putterill *et al.*, 1997; Ledger *et al.*, 1996; Lippuner *et al.*, 1996). Recent Blast comparisons to the GenBank databases (Altschul *et al.*, 1990) have isolated probably fourteen unique *CO* homologues from *Arabidopsis*, which have greatly contributed to the analysis of the conserved amino acids affected in the mutant alleles. In addition, probably nine rice homologues, and one each from maize and pine have been identified (all from EST programmes). Most *CO* homologues identified from genomic sequence appear to contain both the zinc fingers and the conserved C-terminal domain, and it will be interesting to determine how many ESTs, all of which are incomplete, also have both domains and, where sequence information is very limited, whether all have two zinc

fingers. In addition to ESTs, a growing number of homologues are being identified in the genomic sequencing programme, for example, Z97338 on chromosome 4, identified during the sequencing of the ESSA 1 region (European Scientists Sequencing *Arabidopsis*). It will be intriguing to see if any of these mapped homologues correspond to known mutant loci, or flowering time QTLs and whether their patterns of expression might suggest a role in reproductive development.

The molecular analysis of multiple mutations in the *CO* gene identified a number of amino acid residues which are important for the function or stability of domains of the protein. A similar analysis of multiple mutant alleles of the *COP1* gene (*CONSTITUTIVE PHOTOMORPHOGENIC 1*) also identified important domains, but the range of mutations found was in striking contrast to that found at the *CO* locus. The molecular analysis of seventeen recessive *cop1* mutants demonstrated some of the functional implications of the three recognizable structural domains of the COP1 protein, and identified several null mutations (McNellis *et al.*, 1994). Six of the adult-lethal mutants did not accumulate any detectable COP1 protein. The remaining four lethal mutants to be analyzed were compromised in either the G β -protein homology domain or the adjacent C-terminal domain. Only one of these contains a single amino acid substitution; *cop1-9* has a highly conserved glycine residue in the G β -domain substituted by a glutamic acid. The remaining lethal mutants, and the less strongly compromised mutants contain introduced stop codons, either by amino acid substitution or frameshift mutations, or small in-frame deletions or insertions. Interestingly, there were no mutations detected in any alleles in the N-terminal zinc fingers or the adjacent coiled coil region, although the weakest mutant, *cop1-4* contains an in-frame stop codon which results in a truncated protein containing only these two domains. McNellis and colleagues give several explanations for the lack of mutations in the zinc fingers; for example, mutations at the N-terminus of a protein which still retain

a functional C-terminus may be expected to occur at a low frequency because of the likelihood of introducing frameshifts or stop codons early in the protein. The mutants were identified in a screen for recessive constitutive morphogenesis mutants and mutations in the zinc fingers might cause a dominant lethal phenotype. Alternatively they might have a completely different phenotype not detected in their screen.

The contrast between the *cop1* and the *constans* allelic series is striking - six of the seven *co* mutations are single amino acid substitutions, all seven are assumed to be missense mutations affecting structure or function of discrete domains. Five of the mutations affect the N-terminal zinc fingers and do not compromise the remaining C-terminal part of the protein. No mutations introducing stop codons or frameshifts were detected. It is possible that as described for the *cop1* mutants, particular mutations of the *CO* gene, in this case nonsense mutations, would lead to an entirely different phenotype that would not be detected in a screen for late flowering mutants. It would be interesting to know whether the lack of any obvious null mutations at the *co* locus is coincidental and partly caused by the predominance of EMS-induced alleles, or whether a *co* null has a unique phenotype which has so far escaped detection. Chapter 4 describes the screening for, and the identification and characterization of *co-8*, a potential null mutant identified after gamma-irradiation.

4 Isolation and characterization of *co-8*

4.1 Introduction

The seven *co* mutations described in Chapter 3 all caused in-frame changes and none were unambiguous null alleles. In part this may have been a consequence of six of the mutations having been induced with EMS, which usually causes single base changes (Koornneef *et al.*, 1982), however none of the mutations caused premature termination of the CO protein. Furthermore, *co-1*, which was induced with X-rays, was an in-frame deletion of three amino acids. This raised the possibility that all of these alleles retained some CO function, and that perhaps a null allele would cause a more severe effect on flowering time or a different phenotype. All seven *co* alleles were identified in an M₂ screen for late flowering mutants; therefore, if a null allele caused a different phenotype it would not have been selected. It is also possible that the *co* alleles identified so far are dominant because the mutant proteins antagonize the function of the product of a related gene, and that a null allele would cause a wild type phenotype because other genes can compensate for the loss of CO function. In contrast, plants homozygous for a *co* null allele may have a more severe phenotype than those already available, for example never flowering or causing lethality. The semi-dominant *gai* (gibberellic acid insensitive) mutation of *Arabidopsis* confers a dwarf phenotype resembling that of mutants defective in gibberellin biosynthesis (Peng and Harberd, 1993). The *gai* allele was found to be a dominant gain-of-function allele which encodes an altered protein. Derivative alleles of *gai*, which were found to be null alleles, conferred a phenotype indistinguishable from wild type. It was proposed that in the complete absence of a GAI protein a closely related gene such as the highly homologous *GRS* (*GAI*-related sequence) could substitute for its function (Peng *et al.*, 1997). Functional

redundancy has also been proposed for components of the *Arabidopsis* ethylene signal transduction pathway. The alleles isolated at the *ETR1* (ethylene responsive1) locus, *etr1-1* to -4, are all dominant gain-of-function alleles which encode altered proteins. All four are single amino acid substitutions. Again, a closely related gene, *ERS* (ethylene response sensor), is thought to encode overlapping functions which may explain why null alleles with an altered ethylene response have not been isolated for either gene (Bleecker *et al.*, 1988; Chang *et al.*, 1993; Hua *et al.*, 1995).

As well as providing an interesting new allele to study, a null allele of *CO* would be a useful tool for further genetic and molecular analysis. Mutated copies of the CO protein could be introduced into plants that completely lacked CO activity and the effect of the introduced CO proteins could be studied. An allele which does not produce a *CO* transcript would be useful as a negative control in the analysis of *CO* expression by RT-PCR or *in situ* hybridization. The genetic analysis of the relationships between mutations in the same or different flowering pathways would also be simplified by the availability of definite null alleles rather than possible "leaky" or dominant gain-of-function alleles.

At Nottingham University a method for the isolation of deficiencies (deletions) in the *Arabidopsis* genome was developed (Vizir *et al.*, 1994). These experiments were envisaged to provide overlapping deletions or other rearrangements that would enable the integration of physical and genetic maps and the isolation of genes by chromosome walking or genomic subtraction. The technique involved mutagenizing wild type Landsberg *erecta* pollen with gamma-irradiation and using the irradiated pollen to fertilize a line homozygous for a series of recessive mutations which have readily identifiable visible phenotypes (Figure 4.1). If the mutagenized pollen carried a new mutation at one of the loci mutated in the marker line, the phenotype would be visible in the M_1 generation. If these M_1 plants

were then self-fertilized and the next generation screened, all M_2 plants would show the mutant phenotype while the other recessive mutations not affected by the mutagenesis should segregate in a 3:1 wild type:mutant ratio. The material generated from such a screen might provide a useful source of new *co* alleles. One of the marker lines used by Vizir and co-workers was the chromosome 5 marker line *lu co-1 ms1 ttg* (*lu*: centre of the rosette is yellow green, map position 10 cM; *co-1*: late flowering, map position 13 cM; *ms1*: male sterile, map position 23 cM; *ttg*: yellow seeds and no trichomes, map position 28 cM; NASC240; Figure 4.1). As the *co-1* allele is semidominant, the M_1 plants from a cross to wild type Landsberg *erecta* will have a flowering time phenotype intermediate between wild type and a *co* homozygote. Identifying new mutations at the *CO* locus is therefore not as straightforward as identifying new mutations at one of the other loci carrying recessive mutations.

To aid Vizir and co-workers in the identification of individuals carrying new *co* mutations M_2 seeds derived by self-fertilization from late flowering M_1 plants were sent from Nottingham to the Coupland laboratory. The progeny of 26 late flowering M_1 plants potentially carrying new *co* alleles, from the gamma-irradiation screen were kindly supplied by Dr. Igor Vizir.

4.2 Screening for new *co* alleles in an M_2 population made from gamma-irradiated pollen

To identify new alleles at the *CO* locus approximately 20 seeds from each of the 26 self-fertilized M_1 plants were sown in the greenhouse to screen for lines in which all of the M_2 progeny were late flowering. Lines which were late flowering in the M_1 generation because of the dominant effect of *co-1* or of mutations in other flowering time genes or because of environmental effects should segregate early flowering wild type plants in the M_2 . Twenty of the M_2 populations contained early flowering plants (Table 4.1), so work with these lines

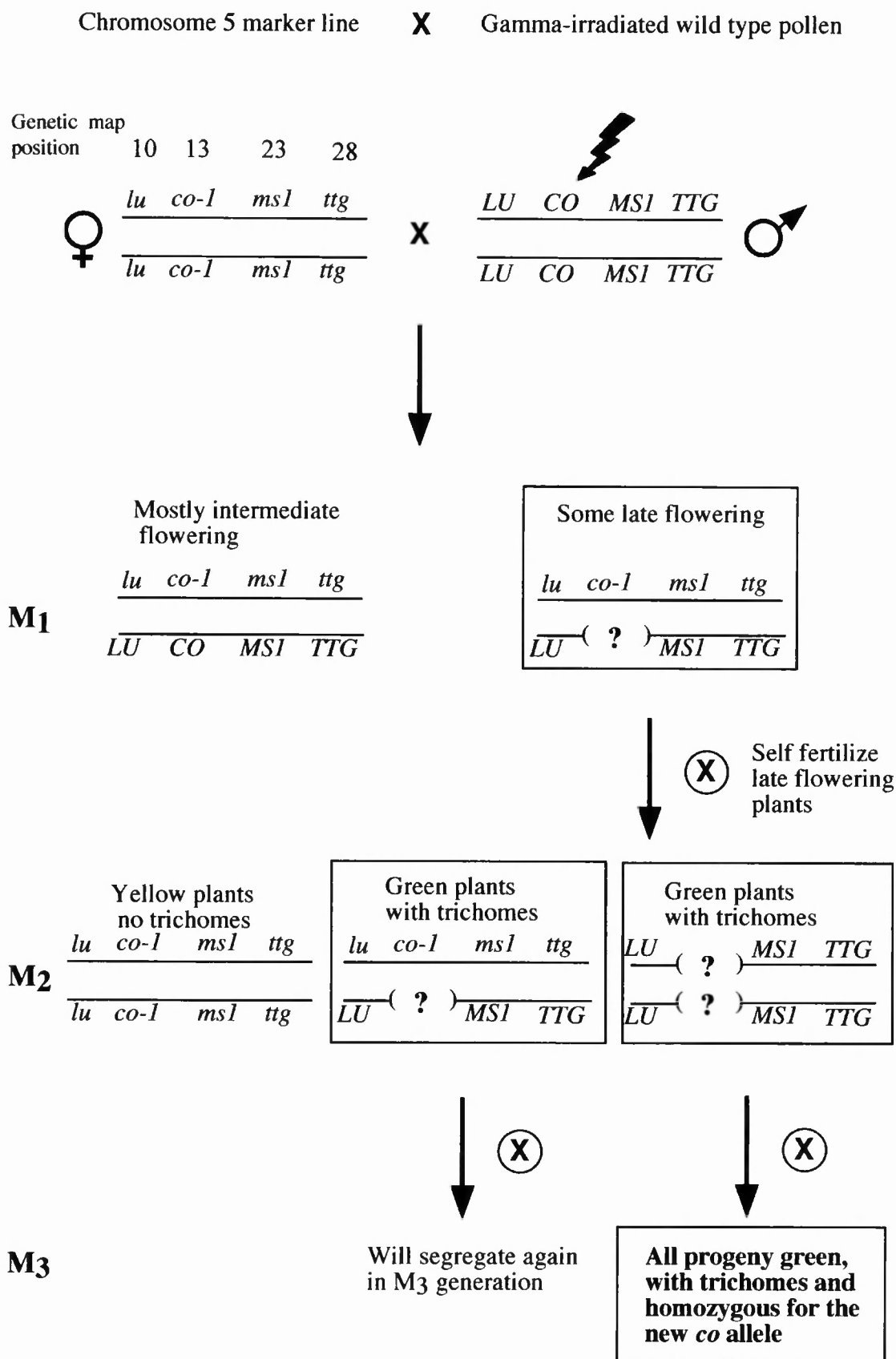


Figure 4.1 The strategy for generating new *co* alleles by gamma-irradiation of wild type pollen. This was then crossed to a line carrying a number of phenotypic markers and potential late flowering individuals identified in the M₁ generation. Seeds from these individuals were provided to the author by Igor Vizir. The genotypes of interest identified in each generation are outlined.

was discontinued. The remaining six lines produced only late flowering progeny. These lines were called co-Dfa2, co-Dfa3, co-Dfa8, co-Dfa9, co-Dfa10 and co-Dfa15 (Table 4.2). The segregation of *lu* was also recorded in the M₂ generation. An increased frequency of *lu* homozygotes in the M₂ above the expected 25% would indicate reduced transmission of the irradiated parental chromosome (Vizir *et al.*, 1994). Five of the six lines: co-Dfa2, co-Dfa3, co-Dfa9, co-Dfa10 and co-Dfa15, showed a distortion in favour of *lu*, which suggests that the new mutant chromosome was poorly transmitted, or in the homozygous state was lethal.

In order to confirm the late flowering phenotype in the M₃ generation and to identify plants homozygous for the new alleles from each of the six lines, five green M₂ plants which were either wild type or heterozygous for *LUTESCENS*, were self-fertilized and M₃ seeds collected. The M₂ plants from lines co-Dfa3 and co-Dfa9 were all found to be sterile so work with these was discontinued. The remaining four lines were analyzed in the M₃ generation (Table 4.2). Line co-Dfa15 segregated early flowering plants in the M₃ generation so work with this line was also discontinued. For the remaining three lines, co-Dfa2, co-Dfa8 and co-Dfa10 all plants from the five families screened in the M₃ generation were late flowering. However, for two lines, co-Dfa2 and co-Dfa8, no *LUTESCENS* homozygotes (100% green M₃ families) were identified, suggesting that perhaps homozygotes of the new *co* alleles in these lines were difficult to recover.

4.2.1 Identification of a new allele, *co-8*

A potential *co* deletion homozygote was identified for line co-Dfa10. M₃ family E325 (Table 4.2) contained only green, fertile, late flowering plants with a normal distribution of trichomes, suggesting that the parental M₂ plant was homozygous for *LUTESCENS*, *TTG* and *MS1* and a new *co* allele. This new allele was named *co-8*.

M₂ line from selfed late flowering M₁ plant	All late flowering in M₂ or segregating earlies?	<i>LU</i> : <i>lu</i> segregation	Comments
Dfx4-1	Segregating earlies	1.5:1	Discontinued analysis
Dfx1	“	nt	“
Dfx1(b)	“	nt	“
Dfx2	“	nt	“
Dfx4	“	nt	“
Dfx5	“	nt	“
Dfx6	“	nt	“
Dfx7	“	nt	“
Dfx7(b)	“	nt	“
Dfx15	“	nt	“
Dfx16	“	nt	“
co-Dfa-1	“	nt	“
co-Dfa-2	LATE	1.86:1	Continued analysis
co-Dfa-3	LATE	0.1:1	Continued analysis
co-Dfa-4	Segregating earlies	nt	Discontinued analysis
co-Dfa-5	“	~3:1	“
co-Dfa-6	“	~3:1	“
co-Dfa-7	“	~2:1	“
co-Dfa-8	LATE	9.0:1	Continued analysis
co-Dfa-9	LATE	0.7:1	Continued analysis
co-Dfa-10	LATE	0.45:1	Continued analysis
co-Dfa-11	Segregating earlies	0.1:1	Discontinued analysis
co-Dfa-12	“	~2:1	“
co-Dfa-13	“	~3:1	“
co-Dfa-14	“	~3:1	“
co-Dfa-15	LATE	1.6:1	Continued analysis

Table 4.1 The screen for lines which do not segregate early flowering plants in the M₂ generation and which therefore may carry a new *co* allele. Each line was derived from a selfed M₁ putative late flowering individual, identified by Igor Vizir and sent to the author for analysis.

Name	M ₂ phenotypes	M ₃ from selfed M ₂ green plants (wild type <i>LUTESCENS</i> homozygotes or heterozygotes)			
		Family	All late?	<i>LU</i> / <i>lu</i>	Comments
co-Dfa-2	All late <i>LU</i> / <i>lu</i> 1.86:1	E312	Yes	3.1:1	No <i>LU</i> homozygotes so no homozygotes for a new <i>co</i> allele
		E313	“	2:1	
		E314	“	2.3:1	
		E315	“	2.6:1	
		E316	“	2:1	
co-Dfa-3	All late <i>LU</i> / <i>lu</i> 0.1:1 but STERILE so discontinued				
co-Dfa-8	All late <i>LU</i> / <i>lu</i> 9:1	E317	Yes	2.6:1	No <i>LU</i> homozygotes so no homozygotes for a new <i>co</i> allele
		E318	“	1.9:1	
		E319	“	2.3:1	
		E320	“	2.6:1	
		E321	“	7:1	
co-Dfa-9	All late <i>LU</i> / <i>lu</i> 0.7:1 but STERILE so discontinued				
co-Dfa-10	All late <i>LU</i> / <i>lu</i> 0.45:1	E322	Yes	nt	Poor germination
		E323	“	4:1	
		E324	“	6:1	
		E325	“	1:0	Homozygous for a new <i>co</i> allele?
		E326	“	4.3:1	
co-Dfa-15	All late <i>LU</i> / <i>lu</i> 1.6:1	E327	NO segregated EARLIES so discontinued		
		E328	NO segregated EARLIES so discontinued		

Table 4.2 The analysis in the M₃ generation of the six late flowering lines picked from the M₂ screen and the identification of a line homozygous for wild type *LUTESCENS* and therefore potentially homozygous for a new *co* allele.

4.3 Characterization of *co-8*

To characterize *co-8*, seeds for further genetic analysis, and plant tissue for DNA extraction were harvested from several individuals in family E325 and the parental M₂ plant. One individual, E325-25, was backcrossed to wild type Landsberg *erecta* to characterize the heterozygote and to segregate *co-8* from the mutagenized background. The genetic analysis of *co-8* is described in section 4.3.2.

4.3.1 Molecular characterization of *co-8*

When *co-8* was identified detailed restriction maps of the region around the *CONSTANS* gene were available and the gene had been cloned and sequenced. It was therefore possible to characterize the *co-8* allele at the molecular level.

4.3.1.1 Analysis of *co-8* by Southern hybridization

The rearrangement in *co-8* was first analyzed by Southern hybridization. DNA extracted from a number of *co-8* M₃ and M₂ individuals, wild type, *co-1* and the two additional potential *co* deletion alleles *co-Dfa3* and *co-Dfa8*, was digested with *Hind* III. After fractionation and Southern blotting the filter was probed sequentially with a number of restriction fragments derived from the region around the *CO* gene. A 2.3 kb *Eco*R I-*Pst* I fragment encompassing most of the first *CO* exon and the promoter region first revealed a rearrangement in *co-8*, and the result of this hybridization is shown in Figure 4.2. Two main observations were made from this preliminary experiment. The first observation was that the hybridization pattern in the individuals carrying a *co-8* allele must be the result of an insertion or inversion at the *CO* locus because of the presence of two additional hybridizing fragments that were larger than those present in wild type and which could not result from a deletion alone. The second was that an individual from the presumed homozygous M₃ family and its M₂ parent contained a wild type hybridization pattern in

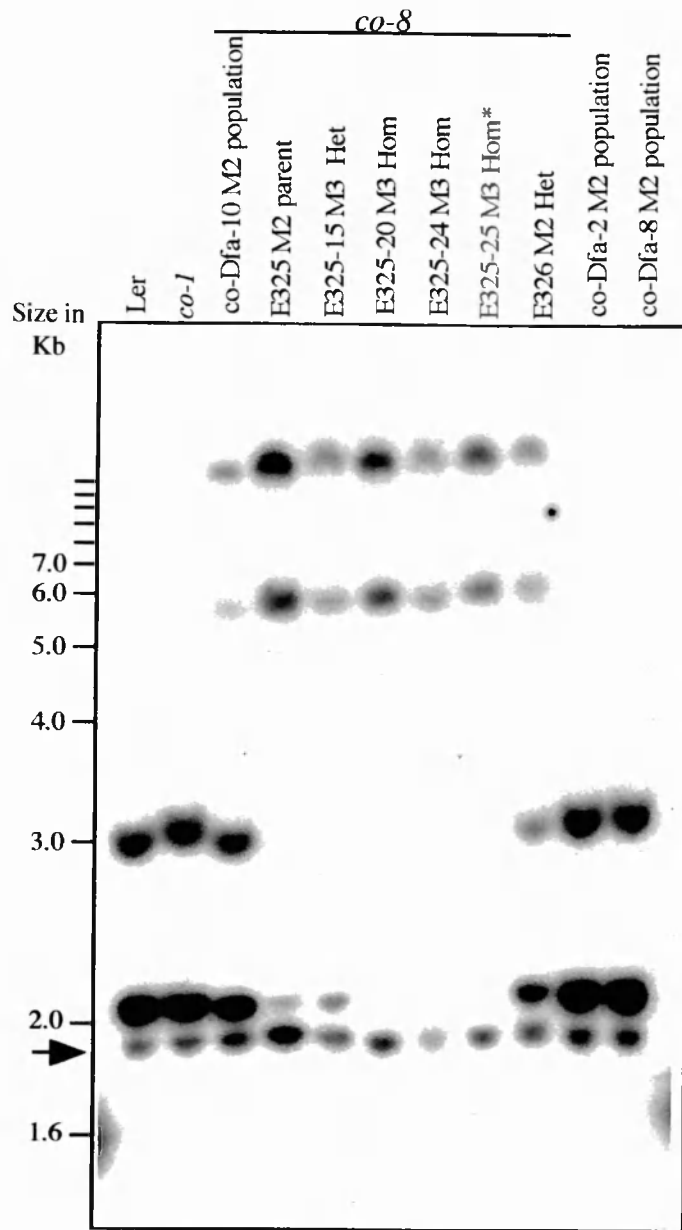


Figure 4.2a Southern blot of plant DNA digested with *Hind* III and probed with the 2.3 kb *Eco*R I-*Pst* I fragment shown below. DNA was extracted from the leaves of individual plants (E325 M2 and M3 and E326 M2 individuals) which were then allowed to self-fertilize. Plant E325-25 (red) was chosen for further genetic analysis. DNA from a population of co-Dfa-10 seedlings grown from the original seed batch sent from Nottingham was included to verify that the *co-8* mutation was found in this line. DNA from populations of co-Dfa-2 and -8 seedlings was also included but no mutations in the *CO* gene in these two lines were identified by Southern analysis. The arrow indicates a hybridizing fragment from a *CO* homologue.

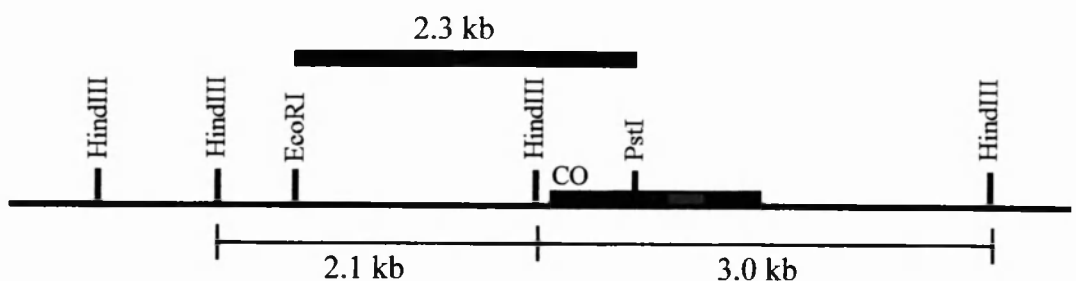


Figure 4.2b Restriction map of the *CO* genomic region with the fragment used to probe the blot above, and the expected wild type *Hind* III hybridizing fragments indicated.

addition to a *co-8* rearrangement. This demonstrated that *lutescens*, *ttg* and *ms1* had been lost from this family, perhaps by a double recombination event on either side of *co-1*, but that the *co-1* allele, which gives a wild type restriction and hybridization pattern, was still present.

The Southern analysis did, however, identify three M₃ individuals which appear to only have a *co-8* hybridization pattern (E325-20, E325-24 and E325-25) and which are assumed to be homozygous for the new allele. DNA from one of these, E325-25, was used in further Southern analysis.

No rearrangements were detected in lines co-Dfa2 or co-Dfa8, with any of the probes used, including the two cosmids 129 and 142 which cover 30 kb around *CO* (data not shown). If these lines contained large deletions it is possible that the irradiated chromosomes were not transmitted through the gametes and were lost, so the lines are homozygous for *co-1* and have lost the flanking markers by recombination. Alternatively the deletions could be very small and therefore only detectable by sequencing the *CO* gene. The molecular analysis of these lines was discontinued.

Using DNA from the chosen *co-8* homozygote, a heterozygote and controls as before, the Southern analysis was repeated with more restriction enzyme digestions and more probes in the region. This allowed the region affected by the rearrangement at the *CO* locus to be defined, and a deletion identified. Figure 4.3a shows a Southern blot probed with a 950 bp *Hind* III - *Bgl* II fragment 5' of the *CO* ORF (Figure 4.3b). This fragment is completely deleted in *co-8*. For simplicity, DNA on either side of the deletion will be referred to as the *CO* ORF side and the *CO* promoter side. The size of the restriction fragments detected in *co-8* in Figure 4.2a were not consistent with *co-8* being caused by a simple deletion, so it seemed likely that an insertion or inversion had occurred between the two breakpoints of the deletion.

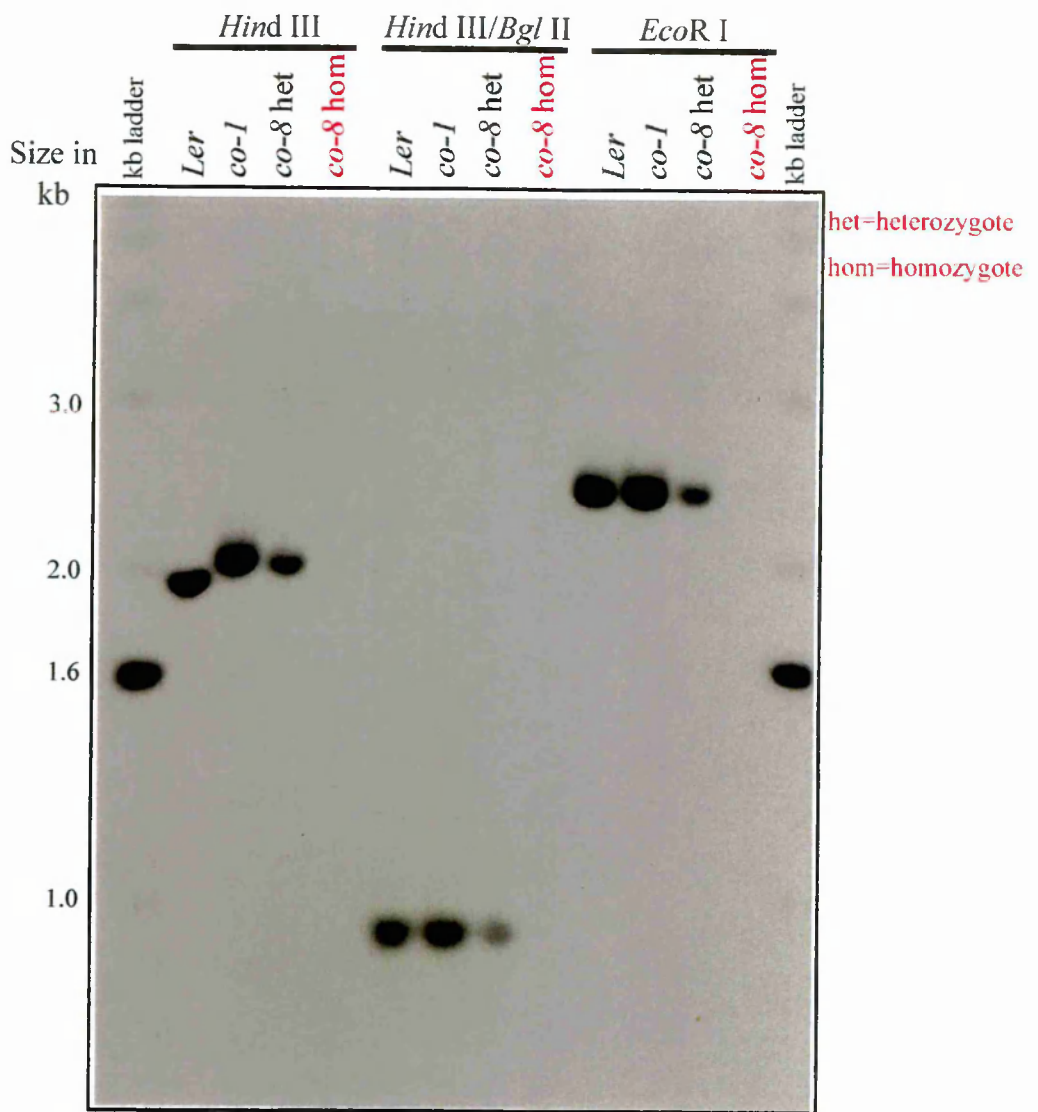


Figure 4.3a Southern blot of plant DNA digested with the restriction enzymes indicated and probed with the 950 bp probe shown below. In *co-8* (in red) this fragment is completely deleted, as the fourth track in each group contains no hybridizing fragment. This blot was reprobed with fragments directly adjacent to the probe used here and fragments not present in wild type were detected in *co-8*, which were not consistent with a simple deletion. Additional Southern blots are not shown as they only confirmed the result already shown in Figure 4.2. The deletion does not extend much beyond this 950 bp region as flanking fragments hybridized to *co-8* DNA.

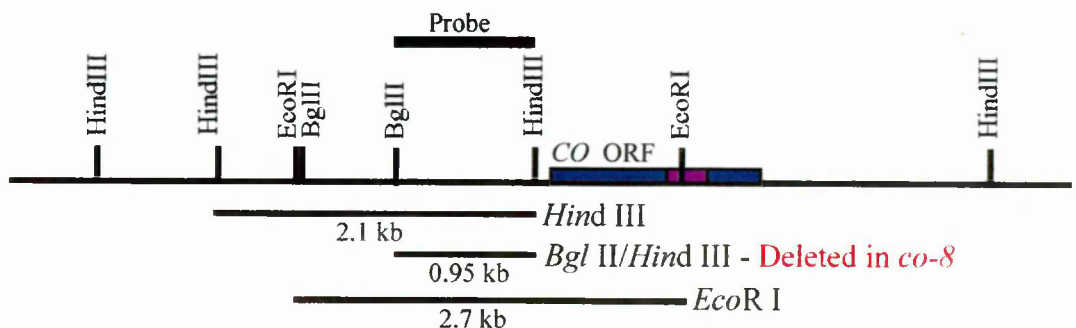


Figure 4.3b Restriction map covering approximately 5 kb in the *CO* region. The fragment used to probe the Southern blot above, which is completely deleted in *co-8*, is indicated, as are the wild type restriction fragments detected by the probe.

Further Southern analysis did not yield any more information about the nature of the rearrangement so it was decided that a cosmid library from the homozygous *co-8* line should be constructed and the DNA introduced at the deletion breakpoints cloned.

4.3.1.2 Construction of a *co-8* cosmid library

Genomic DNA extracted from a population of *co-8* plants was partially digested with *Sau* 3A and ligated into the compatible *Bam*H I site of the cosmid vector 04541. This was packaged into λ -phage which was then used to infect XL1-Blue MR cells. The library was titred and estimated to contain approximately ten haploid genome equivalents.

4.3.1.2.1 Screening the library

The library was amplified and screened simultaneously with two wild type restriction fragments of 1.65 kb and 0.9 kb, one from each side of the deletion in *co-8* (Figure 4.4). After screening 100,000 colonies in two separate experiments seven cosmids were identified but these represented only two unique cosmids, c163 and c164, isolated three and four times, respectively. At the third round of colony purification, probes from each side of the deletion were used separately and it was found that both cosmids were derived from the *CO* promoter side. Subsequent screening of the library with probes from the *CO* ORF side detected no cosmids.

4.3.1.2.2 Characterization of *co-8* cosmid clones by restriction mapping, Southern hybridization and subcloning

Cosmids 163 and 164 were restriction mapped and probed with wild type genomic probes to identify restriction fragments which contained the junction on the *CO* promoter side between DNA flanking the rearrangement and DNA introduced at the deletion breakpoint (summarised in Figure 4.4). As well as a number of recognizable wild type restriction

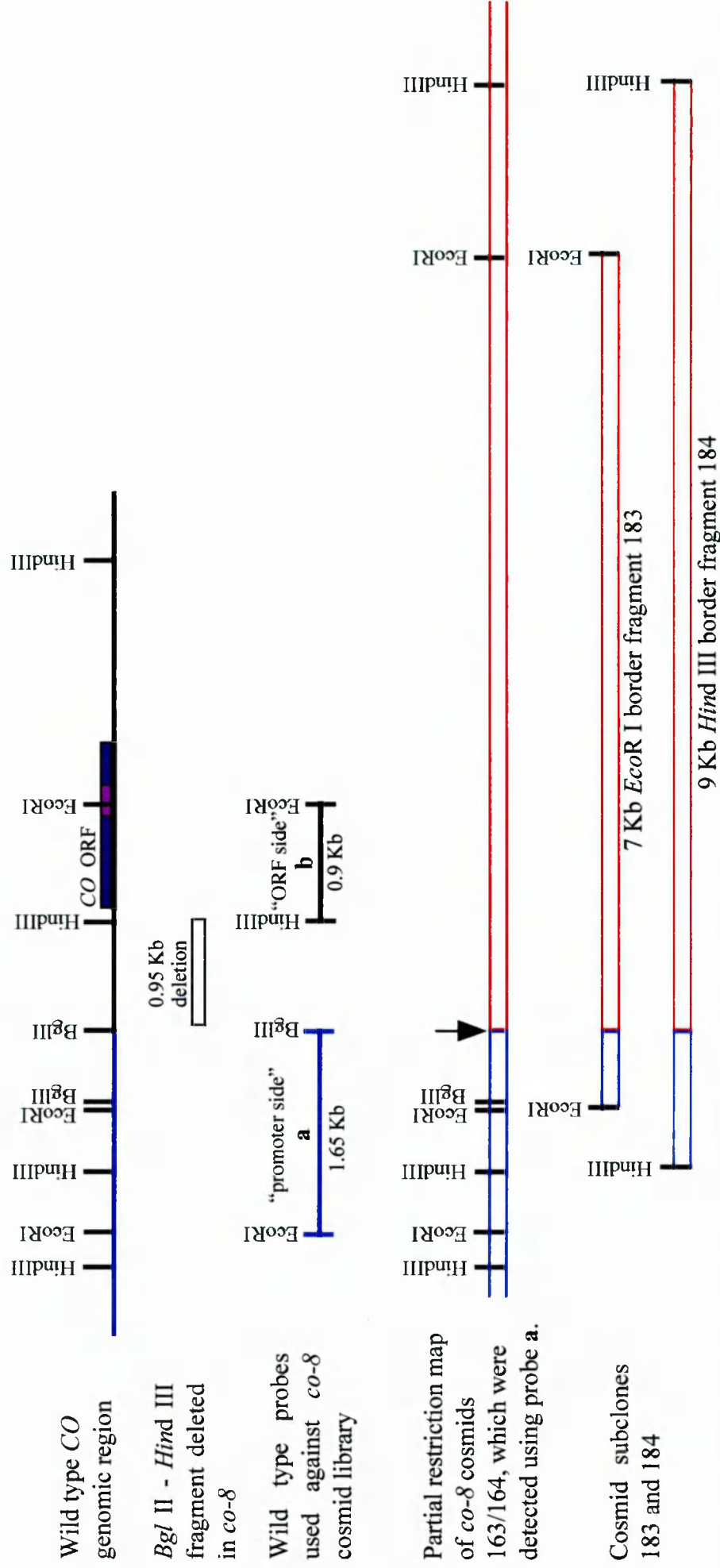


Figure 4.4 The strategy used to isolate DNA adjacent to the *CO* promoter in *co-8*. Fragments **a** and **b** were used as probes against a cosmid library made from *co-8* DNA. Probe **a** (promoter side) detected two cosmids, 163 and 164. Probe **b** (ORF side) detected no cosmids. Cosmids 163 and 164 are both approximately 17 kb and overlap for most of their length. The restriction map of the cosmids is shown for only 10 kb, including approximately 2 kb on the *CO* promoter side of the deletion (in blue) and approximately 8 kb of novel DNA introduced by the *co-8* rearrangement (in dark red). The approximate position of the junction is indicated by an arrow. Large *Eco*R I and *Hind* III fragments not present in wild type but which hybridized to probe **a** must contain the junction and were subcloned for sequencing and mapping.

fragments, these probes identified in both cosmids a 7 kb *EcoR* I fragment, designated 183, and a 9 kb *Hind* III fragment, 184, that are present in *co-8* genomic DNA but not in wild type. These were predicted to contain a junction between the *CO* promoter and a breakpoint in the *co-8* rearrangement. Genomic clones containing the *CO* ORF did not hybridize to these or any of the other cosmid fragments, indicating that the cosmids did not span the rearrangement. To facilitate further analysis the two restriction fragments, 183 and 184, containing the junction between the *CO* promoter and the novel DNA present in *co-8* were subcloned from c163 into pBluescript.

4.3.1.3 Isolation of the DNA flanking the *CO* ORF in *co-8* by IPCR

The *CO* promoter side of the *co-8* rearrangement had been successfully cloned. Cosmids representing the other end of the rearrangement were not detected in the library so an IPCR approach was used to isolate the *co-8* DNA flanking the *CO* ORF. Southern analysis had shown that at least 950 bp of *CO* promoter was deleted but the extent of the deletion in *co-8* into the ORF was unknown. A series of PCR experiments using one primer at the 3' end of the ORF and a number of primers towards the 5' end (shown later in Figure 4.7) revealed that at least 300 bp of the first exon was deleted and that the deleted region included the zinc fingers. This allowed suitable primers to be chosen for Inverse PCR. Figure 4.5 illustrates the IPCR strategy that was used. A Southern blot of *co-8* DNA, digested with a number of restriction enzymes which cut within a suitable distance of the chosen 3' PCR primer was probed with a *CO* clone (data not shown). This revealed that *Bst*Y I cleaved within the rearrangement approximately 150 bp from the breakpoint adjacent to the *CO* ORF. No other enzymes tested produced fragments of a suitable size to be recircularized and amplified by PCR.

To isolate this DNA fragment, *co-8* genomic DNA was digested with *Bst*Y I, self-ligated and used as a template for PCR with primers co25 and co52 (Chapter 7, Materials and

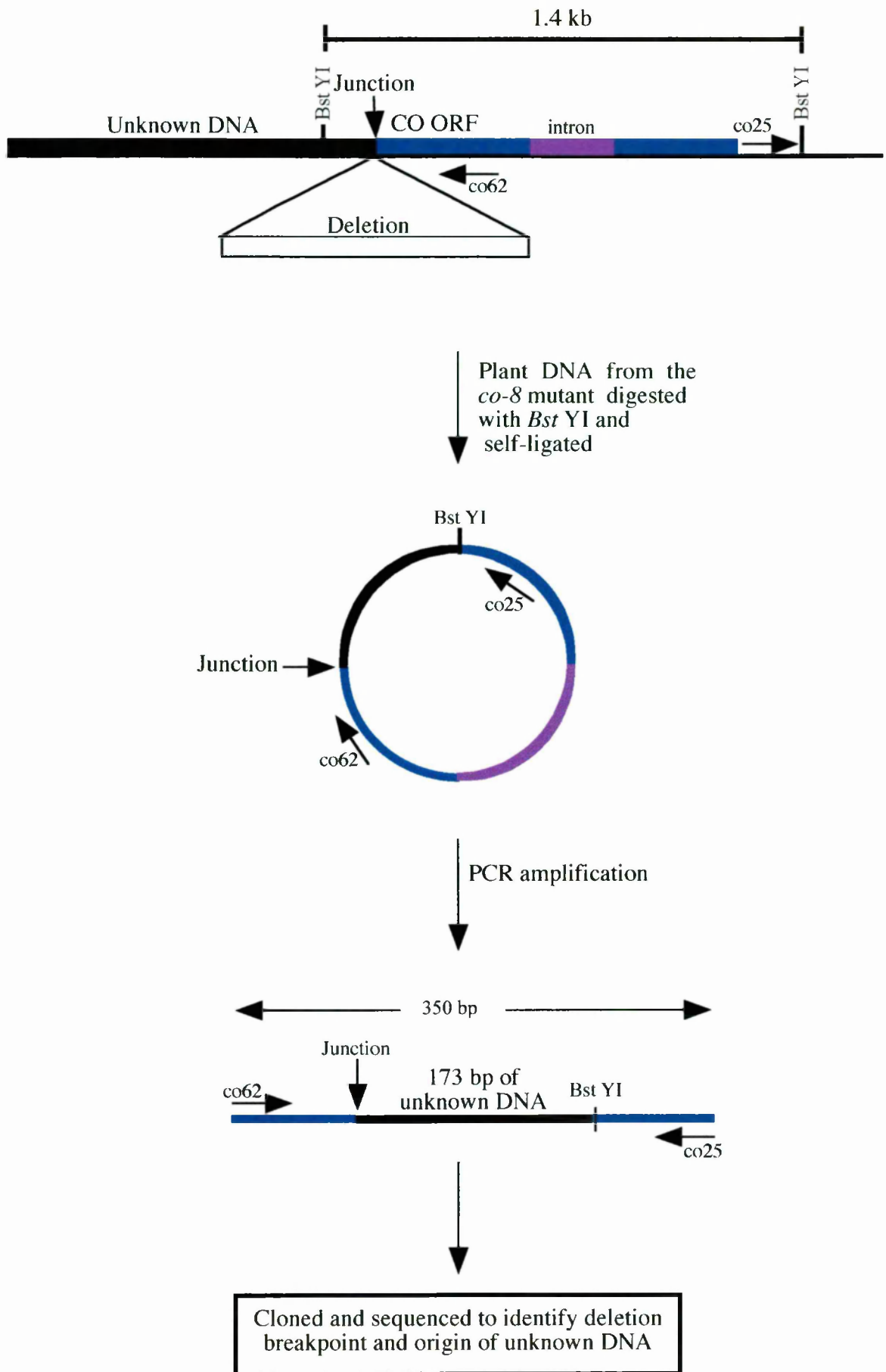


Figure 4.5 Schematic diagram to illustrate how the DNA adjacent to the *CO* ORF in *co-8* was cloned by Inverse PCR. The diagram is not to scale.

Methods) The correct sized PCR fragment, approximately 350 bp, estimated from the Southern blot and the position of the primers, was obtained. After blunt end cloning into pBluescript the PCR product was verified by restriction mapping and sequenced using T3 and T7 primers.

4.3.1.4 The deletion at the *CO* locus in *co-8*

The deletion of at least 950 bp 5' of the *CO* ORF in *co-8* was first detected by Southern analysis of plant genomic DNA. A 950 bp *Hind* III - *Bgl* II restriction fragment failed to hybridize to DNA extracted from a *co-8* homozygote, indicating that the DNA had been deleted and not reinserted elsewhere. Cloning and sequencing the DNA containing each deletion breakpoint allowed the extent of the deletion to be defined. At the *CO* promoter end a suitable primer, *co17*, lay 125 bp upstream of the deleted *Bgl* II site. Using the cosmid subclone p183 as a template, this sequencing primer identified the deletion breakpoint at a position 45 bp upstream of the *Bgl* II site, 1065 bp upstream of the *CO* ATG (Figure 4.6).

At the *CO* ORF side, the 350bp IPCR clone, p12, was completely sequenced using T3 and T7 primers. The cloned border fragment included 173 bp of novel *co-8* DNA between the deletion breakpoint and the *Bst* YI site used for recircularising the IPCR product (Figure 4.5). The breakpoint at this side was 304 bp downstream of the *CO* ATG. Therefore 1369 bp were deleted at the *CO* locus in *co-8*, and this was comprised of 1065 bp extending upstream from the ATG and 304 bp of the first exon (Figure 4.6).

4.3.1.5 Confirmation of the rearrangement in *co-8* by PCR

To verify that the rearrangement in *co-8* was not a simple 1.3 kb deletion, PCR reactions were performed using primers that annealed around and within the deleted region and genomic DNA from wild type and *co-8* mutant plants. Amplified fragments were

TCCTGAAAAT T CCATACAAG TGTGTTTTGT ATTAGATTAC TTCCACAGGT
Bgl II
 TGAGATCTAA TAAAGCTACA ATAAATAGTA TAGAGTATCA TCATAAACCC
 AAATTACAGA GATGTGACAA CACTCATGAG TCATGTTTTG TAACTACTTA
 CTATAATGGT TACCAAGTGC AAATTTCTAC ATACTATATA TGATAAATCT
 AATTATTGCT CATGTGGACT CAAAATGCC TTTTAAGTTT TAACTTGTGC
 GTCAGGTAAA TTCTAATTG TAGTCTCAAG ACTACTTGGC GGATTCGAGT
 TTGATCCTAG AAAATCCACC GTCTCTATGT TTTTCATGTC ACTTTTCCGA
 TATGATTCTC ATTACCATGA CTTTATGAAC CAGATTAAAC ATTATAACAC
 TTTTCATCAG AAAATCCTTC GAAAGTTTCA ATTGCAAATC TTTCTAAATG
 ATGCAGATGC ATTCACAAAT AATGGAACAA CAACTATACC ATATTCACGA
 GTTTGTCTAA CCTTTGTATA GGTAGTCAAC CCATAACAGT TGGTGATGGC
 TCTGACACTC GAAGCCTTAC TCGGAGAGAT ACCTGAACAG TAATCACAAG
 GTTCAGGATG AATATTCAAC CACTTAAACT TTGTATAAAG CCAAAGAGAT
 AAAACGAATC TAGCTTTACT TTAAATAAAA TGCATATGAA AATAGTAAAA
 GGTGATACGA AAAAATAGTA ACAATTTGCC TGCAACACCA TGGCATTATC
 CGGACCACTT CCTCTTGAGA ATCTCAGTAT GGCAAGTGGC AAAACCTAAG
 CAACTTGTGA ACGGGTCCCA ACGAAGAAGT GCATAGGAGG AGATGTTTAC
 ACTTTACACT TTACACTTTA CACTTTACAC ATAGGCCTTC CAAAAGCTC
 AACTAGCTGC AAGAGGATCC AATAACATGT AAGAGCCACT AACGCTGTGC
 CACGTGTAGG CACTCAGGAT TCGATCTTCC CCTCTACTTA TTCTCTCACA
Hind III
 CCAGATATAA GCTT TATTAG CCCCTTCTTT CAGATACCAG CTCCACACC
 ATCAAACCTA CTACATCTGA GTTATT **ATGT** TGAAACAAGA GAGTAACGAC
ATAGGTAGTG GAGAGAACAA CAGGGCACGA CCCTGTGACA CATGCCGGTC
AAACGCCTGC ACCGTGTATT GCCATGCAGA TTCTGCCTAC TTGTGCATGA
GCTGTGATGC TCAAGTTCAC TCTGCCAATC GCGTTGCTTC CCGCCATAAA
CGTGTCCGGG TCTGCGAGTC ATGTGAGCGT GCTCCGGCTG CTTTTTTGTG
TGAGGCAGAT GATGCCTCTC TATGCACAGC CTGTGATTCA GAGGTTTCATT
CTGCAAACCC ACTTGCTAGA CGCCATCAGC GAGTTCCAAT TCTACCAATT
TCTGGAAACT CTTTCAGCTC CATGACCACT ACTCACCACC AAAGCGAGAA

Figure 4.6 The sequence of the DNA which has been deleted from the *CO* locus in the *co-8* allele. The extent of the 1.369 kb deletion is shown in blue. The beginning of the *CO* ORF is underlined, with the ATG start codon in bold. The restriction sites used to make the 950 bp *Bgl* II - *Hind* III probe, which first detected the *co-8* deletion, are shown in purple.

transferred to a filter and probed with a fragment encompassing the *CO* gene, as shown in Figure 4.7. Only primer pairs which lie within the undeleted part of the *CO* ORF enabled amplification of a fragment from *co-8*. Amplification with primers where one of the pair annealed to the *CO* promoter within DNA which, by Southern analysis is still present in *co-8* did not produce amplified fragments. If the rearrangement in *co-8* was a simple deletion then it would be predicted that primers designed to DNA on either side of the deletion would produce fragments smaller than those amplified from wild type DNA. The *CO* promoter must either have been displaced too far from the *CO* ORF by an insertion so that the intervening fragment was too large to amplify, or displaced by an inversion.

To distinguish between these possibilities the sequence of the DNA introduced next to the deletion breakpoints in *co-8* was used to design new primers for PCR (Figure 4.8). If the rearrangement consisted of an inversion between two breakpoints, one at the *CO* locus and one elsewhere on the same chromosome, then two primers designed to *co-8* DNA, one adjacent to the *CO* ORF and one adjacent to the *CO* promoter, would be in the same orientation but at opposite ends of the rearrangement. This primer pair would be predicted to be pointing towards one another in wild type and in a PCR reaction should amplify a fragment from wild type DNA (Figure 4.9). On the other hand, if the rearrangement consisted of an insertion into *CO* then the primers would be in the opposite orientation and pointing away from each other in both *co-8* and wild type and would not be expected to amplify a fragment in a PCR reaction.

Using the primers described, a fragment was amplified from wild type DNA but not from *co-8* which strongly suggested that an inversion had occurred. The fragment amplified from wild type DNA, however, was larger than predicted. From the sequence data and the position of the PCR primers the predicted size was 284 bp. The actual size of the amplified fragment was 1.6 kb. This indicated that an approximately 1.3 kb deletion had occurred at this second breakpoint, in addition to the 1.3 kb deletion within *CO*, and the intervening

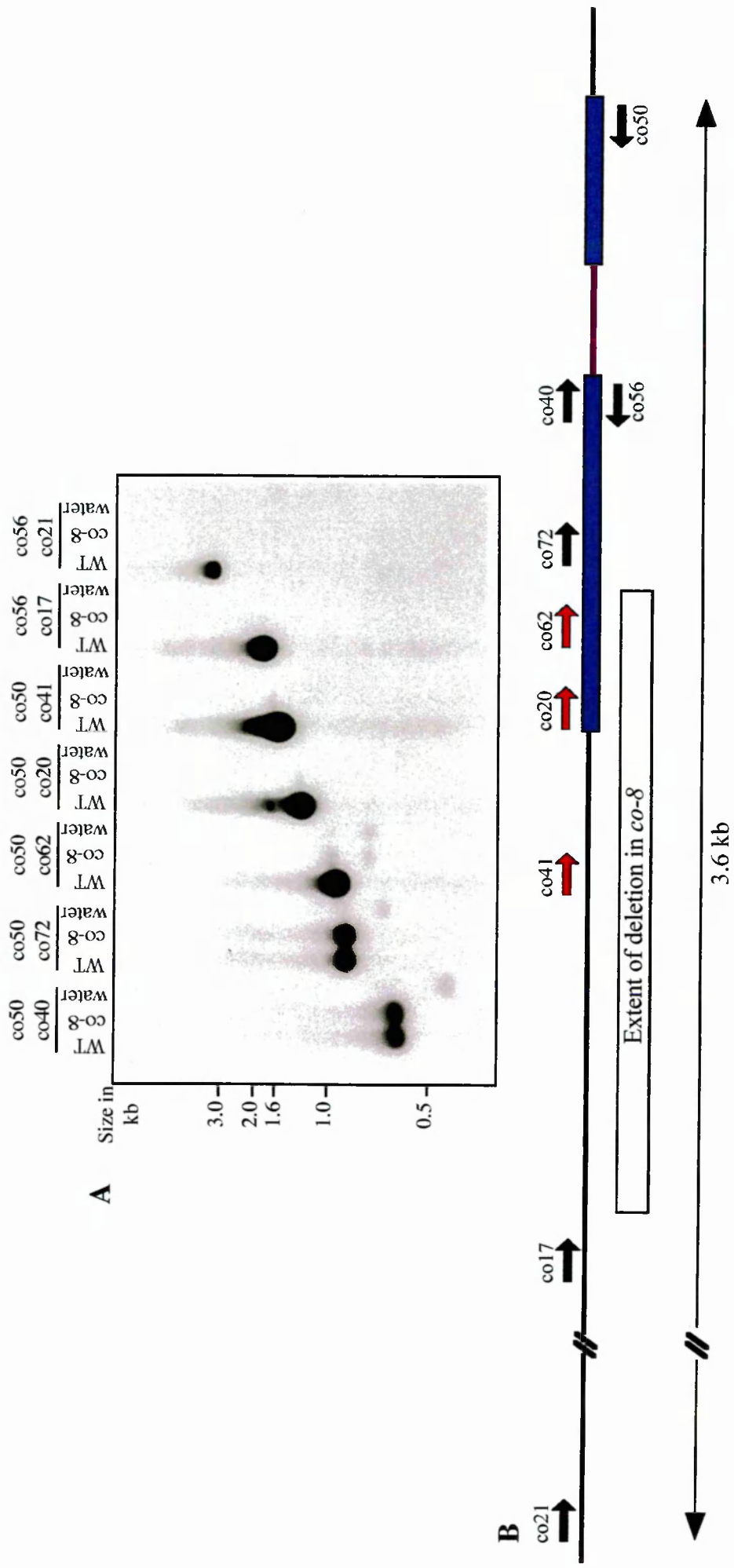


Figure 4.7 Panel A shows a Southern blot of PCR products using the primers indicated and DNA extracted from either wild type or the *co-8* mutant. The blot was probed with the *CO* cDNA. Panel B shows the position of the primers on the wild type *CO* genomic region. The *CO* exons are in blue. No PCR products were amplified from *co-8* DNA using primers co41, co20 or co62 (red) as expected, as these anneal to DNA which is deleted in *co-8*. However, if the rearrangement in *co-8* was a simple deletion, PCR products smaller than those from wild type would be predicted to be amplified from *co-8* DNA using primers co17 and co21. No PCR products were amplified from *co-8* using these primers, indicating that the rearrangement is more complicated.

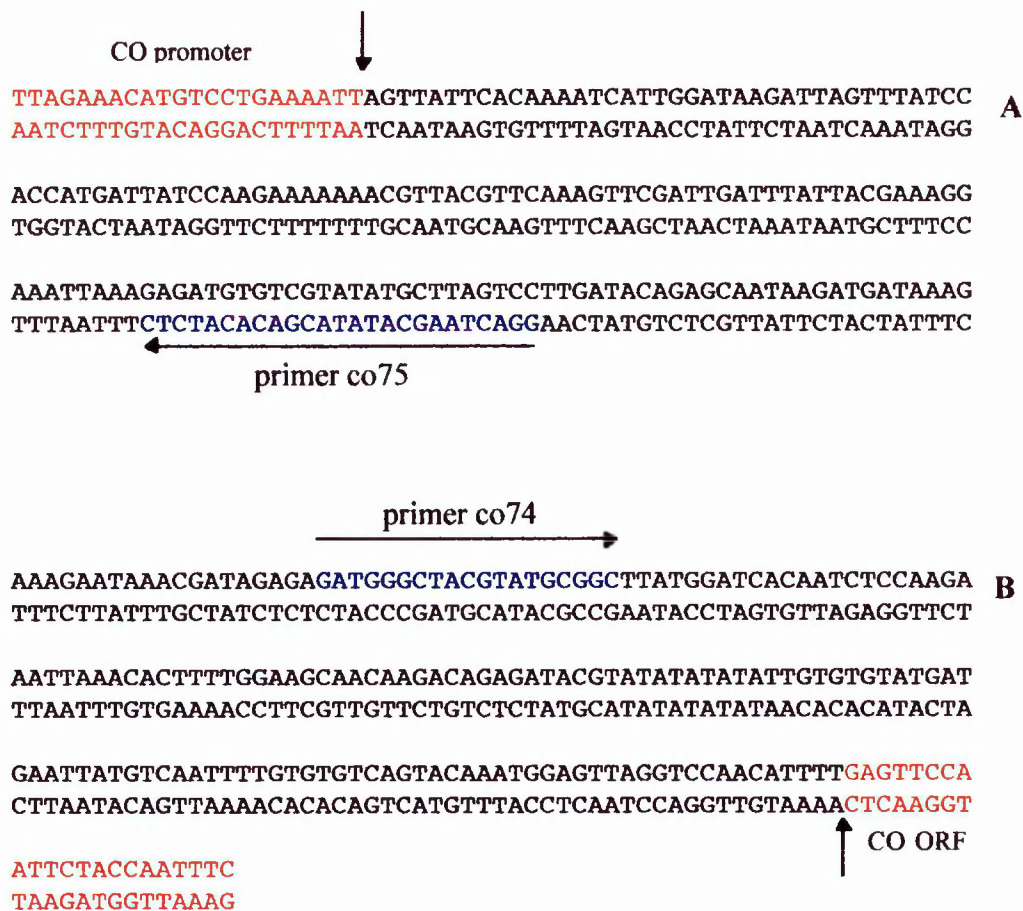


Figure 4.8 DNA sequence of the junctions in the *co-8* allele. The junctions are marked by vertical arrows. The DNA flanking the junctions is red and the DNA introduced at the junctions by the rearrangement is in black. Panel A shows the junction at the *CO* promoter side, and panel B shows the junction on the *CO* ORF side. The primers designed from the sequence introduced at each junction are in blue. These were used to determine whether the rearrangement in *co-8* is likely to be a deletion and insertion, or a deletion and inversion (Figure 4.9).

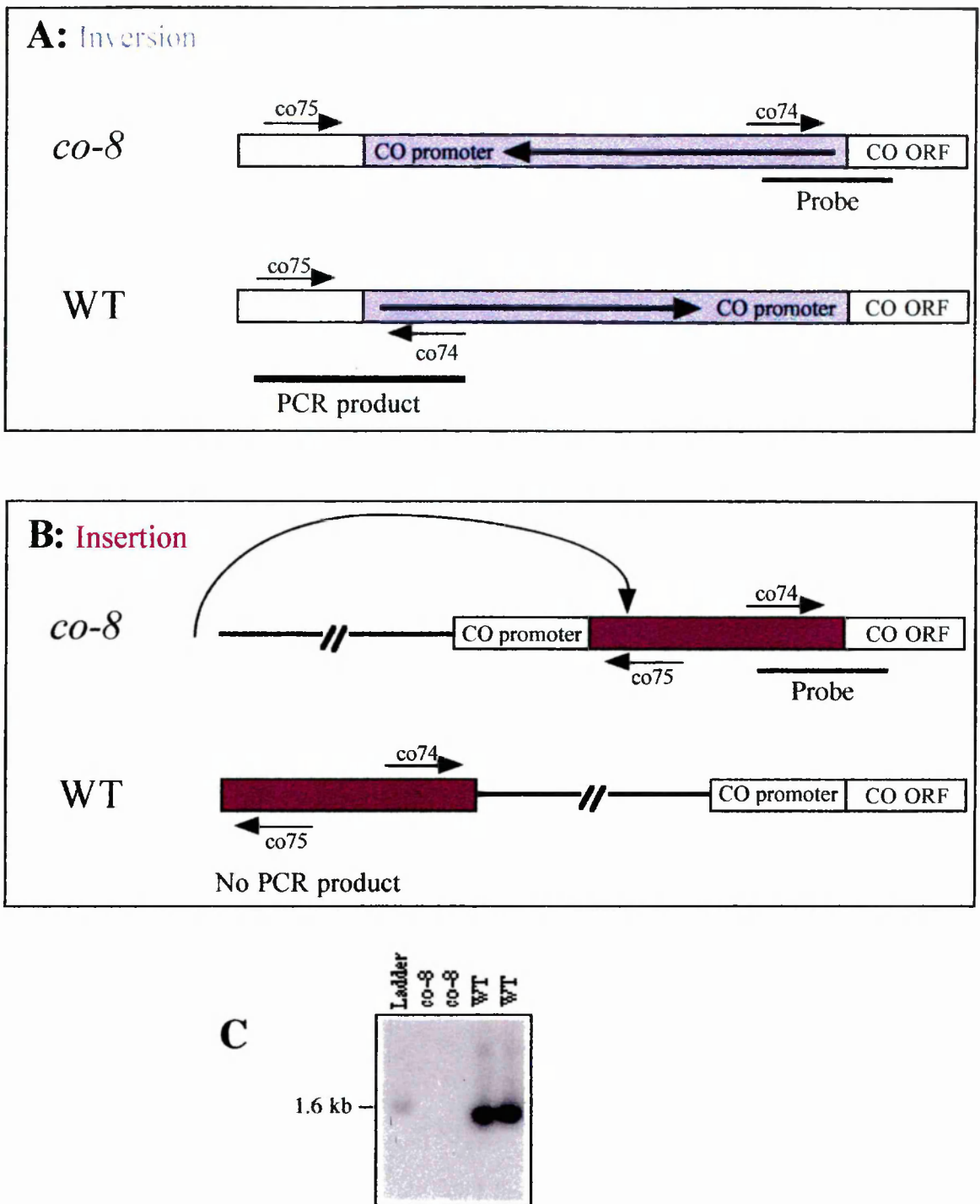


Figure 4.9 The PCR strategy used to establish whether an inversion or insertion had occurred in *co-8*. For clarity, the deletion in *co-8* is not shown and the diagram is not to scale.

Panel A shows the predicted outcome of a PCR reaction with wild type DNA using these primers, if *co-8* contained an inversion. The DNA predicted to be inverted is shown in purple.

Panel B shows that no PCR product would be produced using wild type DNA if *co-8* contained an insertion. The proposed insertion is shown in red.

Panel C shows that a fragment was amplified from wild type DNA but not the *co-8* control samples. This fragment was cloned and sequenced and found to contain recognizable sequence (primers co74 and co75 and sequence up to and including the breakpoints in *co-8*) as predicted in Figure 4.8. The Southern blot was probed with the fragment indicated in Panels A and B.

DNA had been inverted. The 1.6 kb wild type fragment containing the DNA deleted at the second breakpoint in *co-8* was blunt ended and ligated into pBluescript for sequencing.

4.3.1.6 Determining the chromosomal location of the second breakpoint in *co-8*

The availability of anchored YAC contigs generated during the *Arabidopsis* physical mapping programme has provided a means to map DNA clones of interest onto the *Arabidopsis* chromosomes. If the DNA adjacent to the *CO* deletion in *co-8* was represented on a physically mapped YAC or BAC this would eradicate the need for more complicated mapping procedures.

The CIC YAC library (Creusot *et al.*, 1995), consisting of 1152 clones with an average insert size of 420 kb, is estimated to cover at least 92% of the *Arabidopsis* genome. Many of these have now been mapped. Two duplicate filters each containing two copies of the entire CIC YAC library, processed for colony hybridisation were kindly provided by Dr. Melanie Stammers. These were probed with the 9 kb *Hind* III cosmid subclone, p184, containing *CO* promoter DNA and the *co-8* junction. Six YACs were detected, all represented twice on each filter, in a recognisable pattern as would be expected for the CIC filter grid. Four of the YACs (CIC8H6, CIC3F11, CIC8E12 and CIC8H5) contained the *CO* gene and confirmed that at least part of the cosmid subclone used was derived from this region. On the physical map these YACs map to position 33.2 on chromosome 5. The remaining two YACs, CIC9E2 and CIC9F3, mapped to position cM 115.2 on the bottom arm of chromosome 5, and contain markers mi69, mi70, mi418 and m558A (Schmidt *et al.*, 1997; Figure 4.10). The hybridization result was confirmed by demonstrating that the same probe hybridized to the two YACs when they were digested with restriction enzymes and Southern blotted, but not to YACs derived from a different region of the genome that were used as negative controls (Figure 4.11). The YAC Southern blot was reprobed using

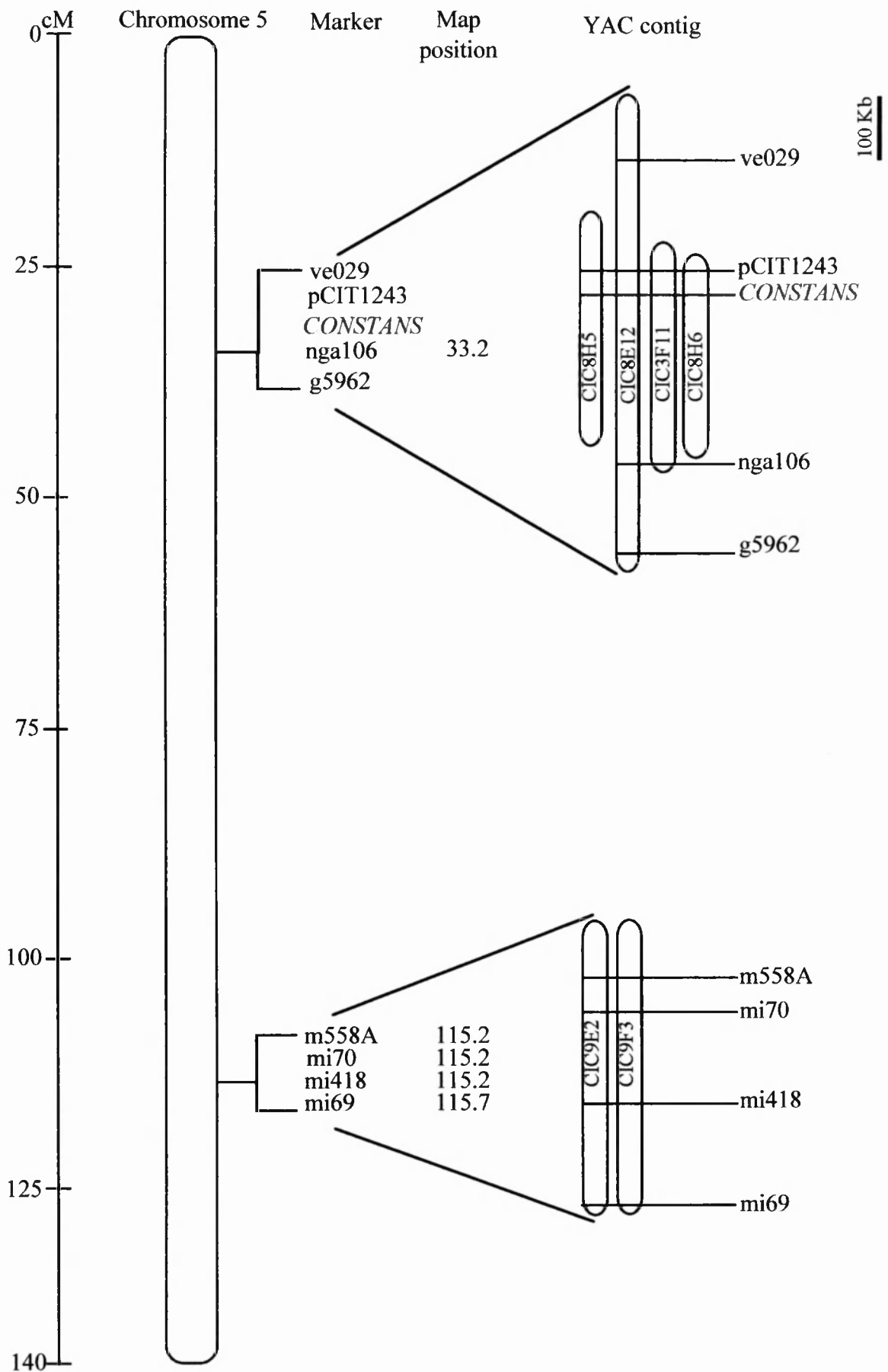


Figure 4.10 Genetic map positions of the YAC contigs at the *CO* locus and the contig at the second breakpoint in *co-8*. The markers were located onto YACs and mapped using recombinant inbred lines by Schmidt *et al.* (1997). The genetic distance between these two contigs is approximately 82 cM, a physical distance of approximately 16.4 Mb

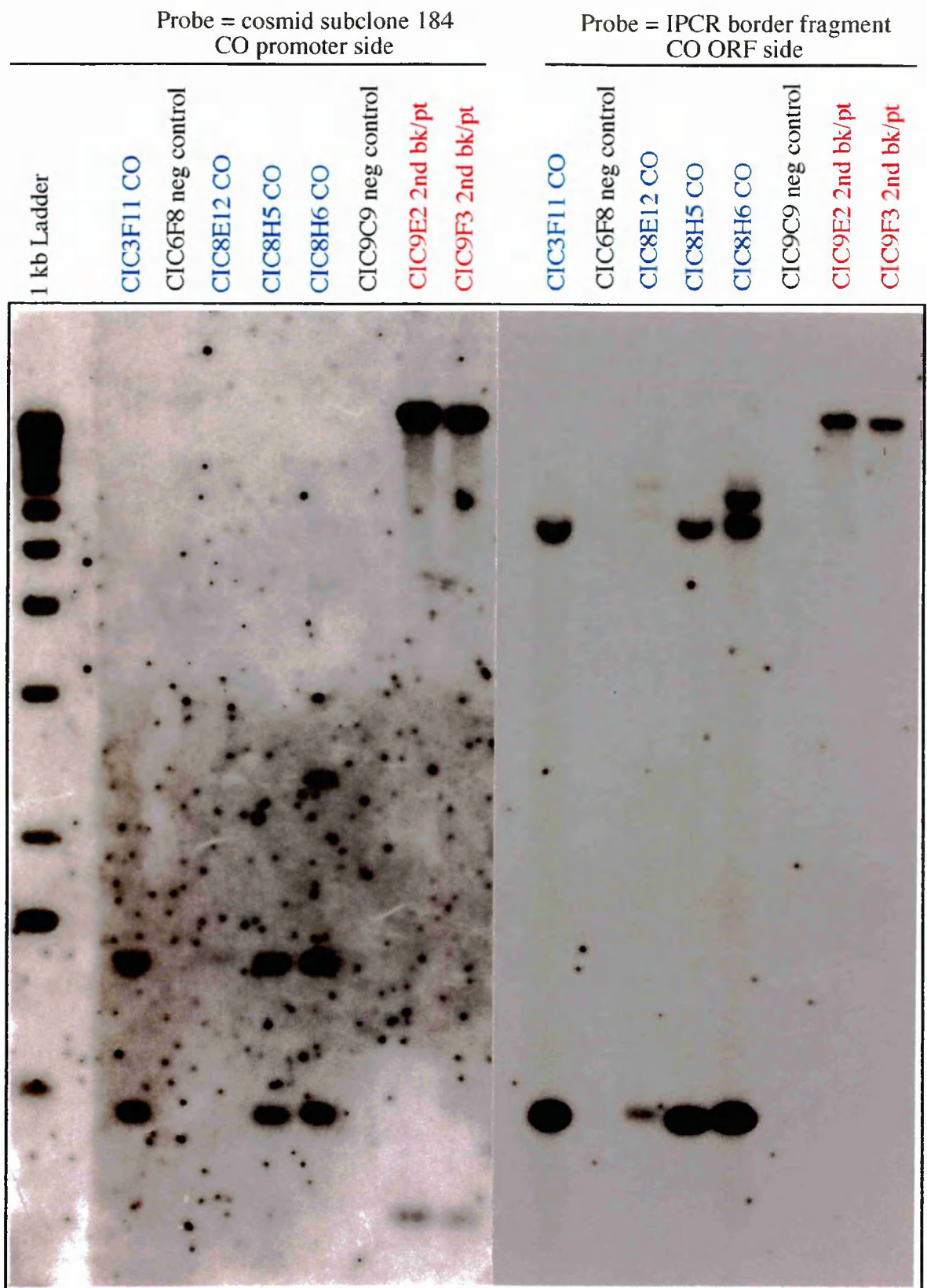


Figure 4.11 Southern blot of YAC DNA digested with *EcoR* I and *BamH* I. The blot was probed sequentially with the two border fragments from either side of the rearrangement in *co-8*, as indicated. The *CO* YACs are in blue. The YACs which map 82 cM from *CO*, where the second breakpoint in *co-8* lies, are in red. The negative controls are in black.

the cloned fragment from the *CO* ORF end of the rearrangement. This probe also hybridized to the four YACs which contained *CO*, and the two which contained the second breakpoint in *co-8*.

4.3.1.7 Analysis of RNA produced by the *co-8* allele

A fragment containing 1.1 kb of upstream sequence and the *CO* ORFs was sufficient to complement the *co-2* mutation and restore wild type flowering time (Kate Dixon, unpublished results). The 1.1 kb region presumably contains all or most of the promoter required for *CO* expression, and is almost completely deleted in *co-8*. To test whether no *CO* mRNA was present in *co-8* plants, RT-PCR was used (Figure 4.12). No transcript was amplified from *co-8* cDNA when one of the PCR primers annealed within the deleted region. Wild type cDNA yielded a transcript of the expected size, which hybridized to a *CO* probe. However, when PCR primers were used that annealed to part of the *CO* ORF that is retained in *co-8*, a hybridizing product of the expected size was amplified from both *co-8* and wild type cDNA. Therefore, although the promoter is deleted in *co-8*, the remaining portion of the *CO* ORF is still transcribed.

4.3.1.8 Sequence analysis of the *co-8* rearrangement

The deletion in *co-8* removed the DNA encoding the first 102 amino acids of the CO protein. This region includes both zinc fingers and part of the basic, conserved domain to the carboxyl side of the zinc fingers (Figure 4.13a). The deletion breakpoint and the sequence introduced by the inversion created an in-frame stop codon that immediately preceded the first amino acid codon that was retained (Figure 4.13b). The 5' portion of the novel transcript produced by *co-8* was not sequenced. However, the presence of an in-frame stop codon immediately prior to the remaining ORF suggested that even if the transcript is translated, a fusion protein containing a novel sequence fused to the carboxyl end of CO

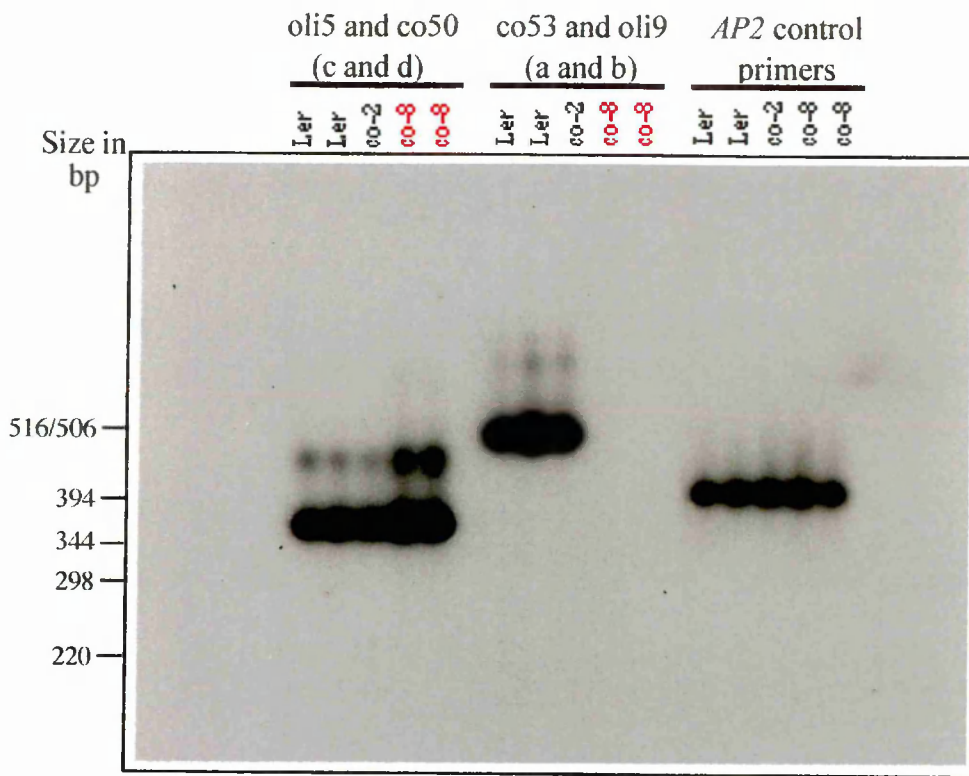


Figure 4.12a Southern blots of RT-PCR products probed with the *CO* cDNA, where amplification was with *CO* primers or with the *APETALA2* cDNA, where amplification was with the *AP2* primers. Duplicate samples (*Ler* and *co-8*) were derived from independent RNA preps and reverse transcription reactions.

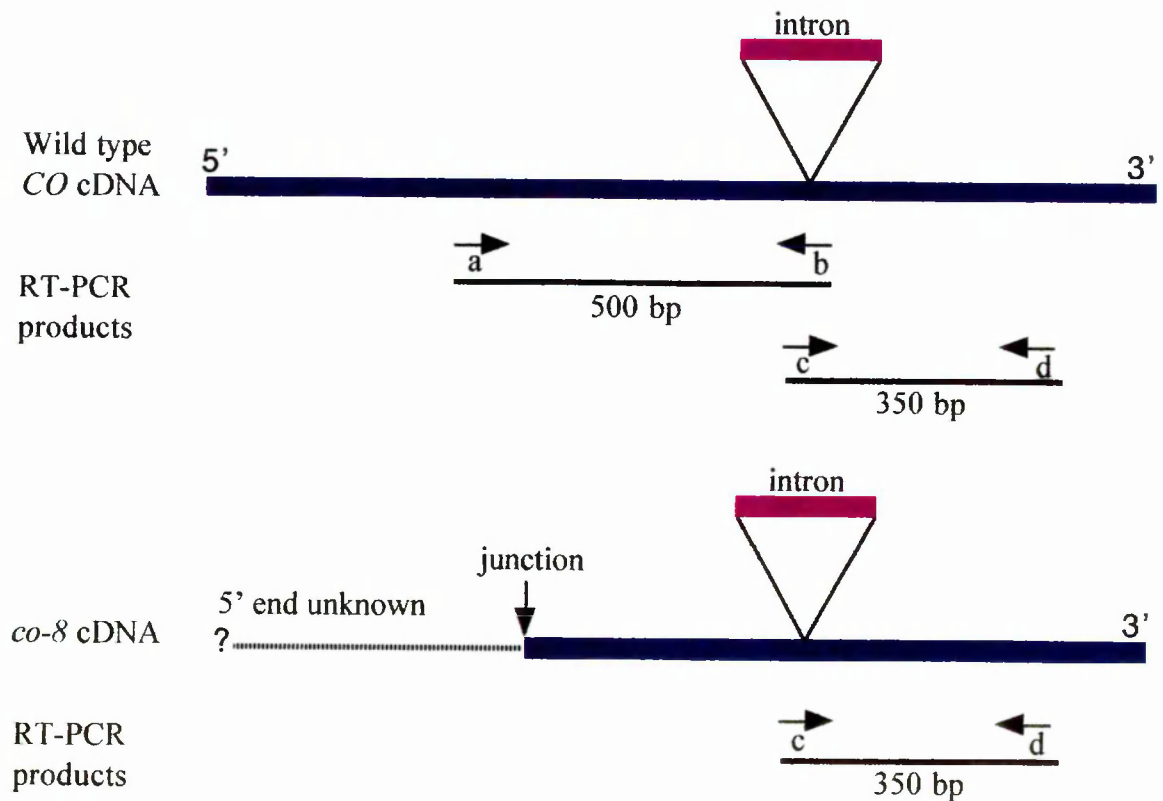


Figure 4.12b The RT-PCR strategy for analyzing *CO* expression in the *co-8* allele. Primers b and c were designed to exon sequence on either side of the intron to prevent amplification of contaminating DNA. Details of primers can be found in Materials and Methods. Primer a = co53, Primer b = oli9, Primer c = oli5, Primer d = co50

is unlikely to be made. There are seven internal methionine codons in the remaining undeleted *CO* ORF sequence but whether *co-8* generates a truncated *CO* protein whose translation is initiated from one of these is not known.

The sequence introduced adjacent to both deletion breakpoints in *CO* was determined and as described previously, PCR primers were designed to this sequence and used to amplify 1.6 kb of wild type DNA from the lower arm of chromosome 5 that was disrupted by the rearrangement. The 1.3 kb "deletion" was sequenced and analyzed to determine whether a second gene as well as *CO* was disrupted in *co-8*. No obvious ORFs were found. Databases were searched for homology to the sequence but no significant homologies were detected. The CIC YACs and BACs from this region have not yet been sequenced. (*Arabidopsis thaliana* DataBase (AAtDB)). Two ESTs have been placed on CIC9E2 and CIC9F3, but the sequence available for both does not correspond to the sequence obtained from the second breakpoint.

4.3.1.9 A summary of the rearrangement in the *co-8* allele

At the *CO* locus a 1.369 kb deletion removed 304 bp from the 5' end of the first *CO* exon plus approximately 1 kb of the 5' untranslated and promoter sequence. On the bottom arm of chromosome 5 a second deletion occurred that removed approximately 1.3 kb, but whether this disrupted another gene is not known. The intervening 82 cM, a physical distance of approximately 16.4 Mb, was inverted. Although the *CO* promoter was deleted, a *CO* transcript was still formed presumably because the DNA brought adjacent to the *CO* ORF by the inversion contained promoter sequences. Whether this truncated transcript could be translated is not known. The *co-8* rearrangement is summarised in Figure 4.14.

MLKQESNDIG SGENNRARPC DTCRSNACTV YCHADSAYLC MSCDAOVHSA
NRVASRHKRV RVCESCERAP AAFLCEADDA SLCTACDSEV HSANPLARRH
ORVPILPISG NSFSSMTTTH HQSEKTM TDP EKRLVVDQEE GEEGDKDAKE
 VASWLFPSND KNNNNQNNGL LFSDEYLN LV DYNSSMDYKF TGEYSQHQQN
 CSVPQTSYGG DRVVPLKLEE SRGHQCHNQQ NFQFN IKYGS SGTHYNDNGS
 INHNAYISSM ETGVVPESTA CVTTASHPRT PKGTVEQQPD PASQMITVTQ
 LSPMDREARV LRYREKRKTR KFEKTIRYAS RKAYAEIRPR VNGRFAKREI
 EAEEQGFNTM LMYNTGYGIV PSF

Figure 4.13a The CO amino acid sequence showing in red the amino acids which are predicted to be deleted from the ORF in *co-8*. The amino acids for which coding sequence still remains in *co-8* are shown in blue. The putative DNA binding domain, for which the coding sequence is almost completely deleted, is in bold and is underlined

KE*TIE
 RWATYAAYGS QSPRN*TLLE ATRQRYVYIY CVYDELCQFC VSVQMELGPT
F*VPILIPSG NSFSSMTTTH HQSEKTM TDP EKRLVVDQEE GEEGDKDAKE
 VASWLFPSND KNNNNQNNGL LFSDEYLN LV DYNSSMDYKF TGEYSQHQQN
 CSVPQTSYGG DRVVPLKLEE SRGHQCHNQQ NFQFN IKYGS SGTHYNDNGS
 INHNAYISSM ETGVVPESTA CVTTASHPRT PKGTVEQQPD PASQMITVTQ
 LSPMDREARV LRYREKRKTR KFEKTIRYAS RKAYAEIRPR VNGRFAKREI
 EAEEQGFNTM LMYNTGYGIV PSF

Figure 4.13b The predicted amino acid sequence of the DNA brought adjacent to CO in the *co-8* allele. The CO amino acids for which coding sequence is still present in *co-8* are shown in blue. The novel sequence brought adjacent to the remainder of CO is in red. There is an in-frame stop codon directly adjacent to the undeleted part of CO suggesting that, even if a stable transcript is made in the *co-8* mutant, it is unlikely that a novel fusion protein will be made.

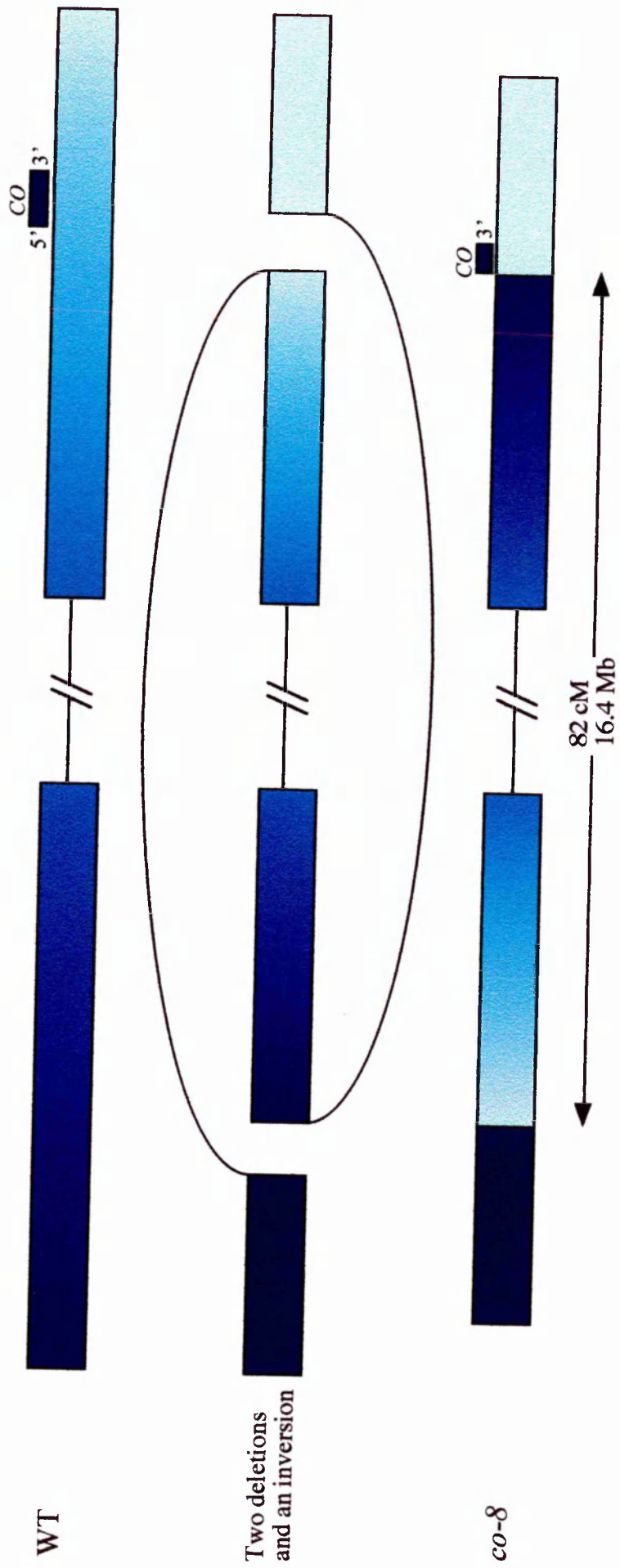


Figure 4.14 The chromosomal rearrangement in *co-8*. The diagram is not to scale.

4.3.2 Phenotypic characterization of *co-8* mutants

A plant homozygous for *co-8*, E325-25, was identified by Southern analysis. This plant was self-fertilized and the seeds collected for phenotypic analysis. One or two flowers from this individual were also fertilized with pollen from a wild type Landsberg *erecta* plant to examine the phenotype of *co-8* heterozygotes.

4.3.2.1 Flowering time

The *co-8* allele was first identified in the M_1 of a cross of mutagenized wild type pollen to a *co-1* homozygote. It was selected as a potential new *co* allele because the M_1 plant was later flowering than its *co-1* heterozygous siblings. Plants homozygous for the *co-8* allele were identified in the M_3 generation and were grown in a controlled environment cabinet in a long day photoperiod (LD - 16 h light, 8 h dark) to measure their flowering time. All eight *co* mutants and a wild type control were sown at the same time, and in the same controlled environment cabinet so that their flowering times could be directly compared. The data for the seven previously isolated mutants were presented in Chapter 3.

Although the *co-8* allele contained a deletion that removed both zinc fingers and is unlikely to retain any *CO* function, other *co* alleles caused later flowering than *co-8* (Figures 4.15 and 4.16; Table 4.3). The other "intermediate" mutants, *co-2*, *co-6* and *co-1*, showed very similar flowering times and their relative strength varied depending on environmental conditions. In this analysis the *co-8* phenotype was intermediate between *co-2* and *co-6*. It was induced to flower after 32.1 days with a total leaf number of 22.5. The wild type Landsberg *erecta* control was induced after 18.9 days with a total leaf number of 7.5 (Table 4.3).

	Flowering time		Leaf number		
Mutant	Days to flowering	Days to first open flower	Rosette	Cauline	Total
<i>Ler</i>	18.9±0.9	27.3±0.8	4.9±0.2	2.6±0.2	7.5±0.2
<i>co-4</i>	24.3±0.5	33.5±0.4	7.6±0.2	3.8±0.1	11.3±0.4
<i>co-5</i>	28.2±0.5	36.4±0.6	9.2±0.2	5.2±0.2	14.4±0.4
<i>co-8</i>	32.1±0.5	43.2±0.5	13.7±0.2	8.8±0.2	22.5±0.3
<i>co-2</i>	33.1±0.9	43.6±0.9	13.5±0.7	5.9±0.3	19.4±0.9
<i>co-6</i>	34.2±0.5	48.7±0.7	13.4±0.3	10.3±0.3	23.7±0.5
<i>co-1</i> Landsberg <i>erecta</i>	36.0±0.6	50.0±0.4	16.1±0.4	9.4±0.3	25.6±0.6
<i>co-7</i>	37.8±0.7	47.5±0.6	18.5±0.5	9.9±0.4	28.5±0.8
<i>co-1</i> Landsberg	38.6±0.8	50.3±0.5	17.9±0.3	12.3±0.2	30.2±0.4
<i>co-3</i>	39.9±0.6	53.4±0.7	17.7±0.8	11.4±0.5	29.1±1.2

Table 4.3 Flowering times and leaf numbers of the eight *co* mutants grown in LDs compared to wild type Landsberg *erecta* . The mutants are listed in order of increasing number of days to flowering. The new *co-8* mutant is highlighted in purple.



Figure 4.15 The *co* mutants grown in LDs, as in Figure 2.4. The new *co-8* mutant has been added (bottom right). All plants are 42 days old.

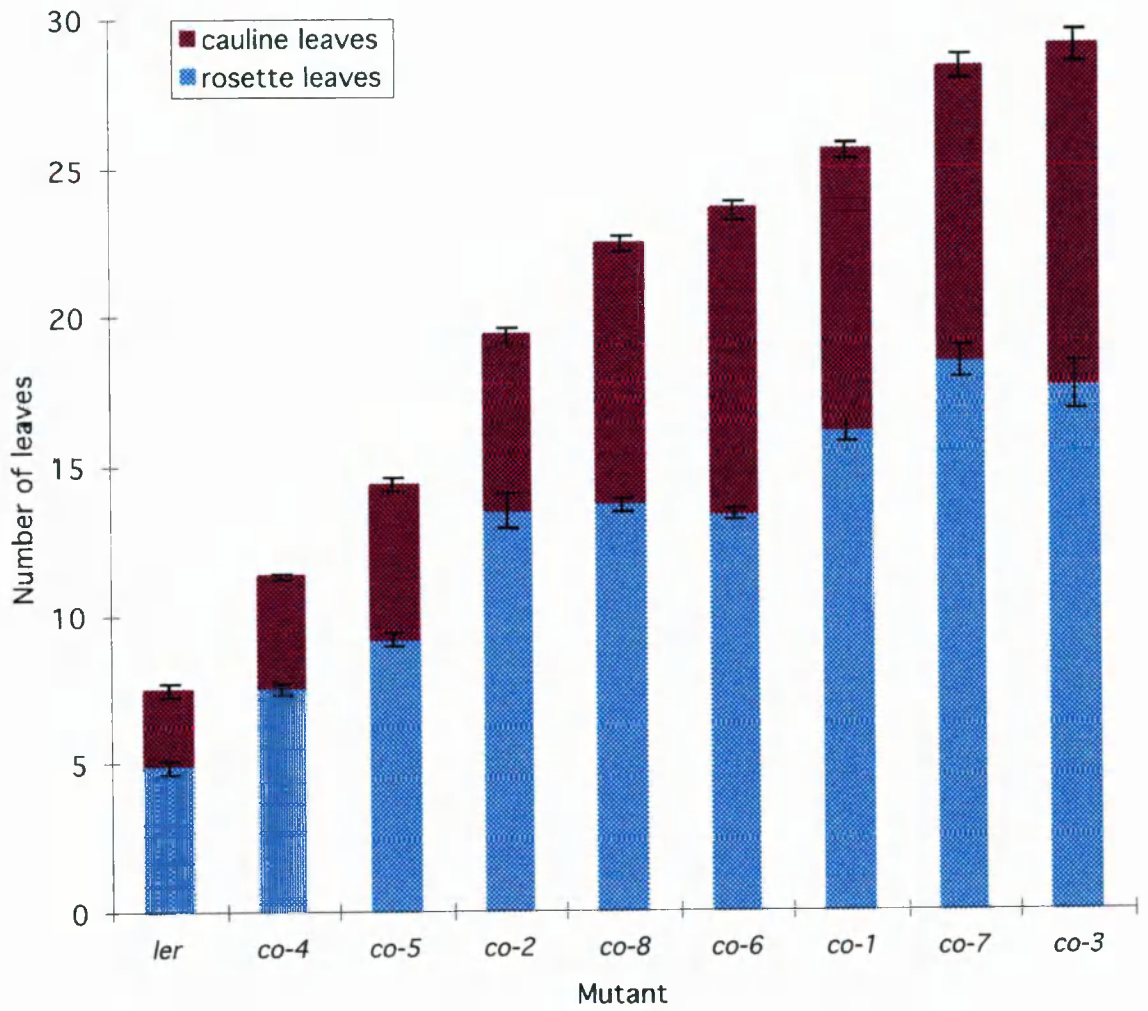


Figure 4.16 The total leaf number of the eight *co* mutants compared to wild type *Landsberg erecta* in LDs. The leaves are divided into rosette and cauline. Standard error bars are in black.

4.3.2.2 Other phenotypes

Although *co-8* plants are similar to mutants homozygous for the other *co* alleles they have a number of additional characteristics which may or may not be caused by the mutation at the *CO* locus. The *co-8* mutant has, compared to the other *co* mutants, a higher leaf number at the same flowering time (Figure 4.17). It also bolts more rapidly than other *co* mutants with similar flowering times (Figure 4.15). A difference in development of plants homozygous for different *co* alleles was observed previously (Koornneef *et al.*, 1991). For example, *co-6* was intermediate in flowering time and leaf number but, in contrast to *co-8*, was delayed in the growth of the bolt and in the opening of its flowers. These effects may be allele-specific.

The first few flowers of a *co-8* plant often open prematurely, before the bolt has started to expand and the petals of these flowers do not completely expand.

Although in *co-8* the ratio of rosette to cauline leaves, and therefore the number of secondary inflorescences, is comparable with the other mutants, the number of cauline leaves (and hence the number of tertiary inflorescences) on the secondary inflorescences seems to be higher although this aspect of the phenotype has not been quantified.

Finally, the cauline leaves of *co-8* often curl upwards and inwards while the plant is first bolting, partly obscuring the primary inflorescence.

Further rounds of backcrossing may remove some of the phenotypes if they are caused by mutations other than *co-8*. Alternatively, introducing a construct containing only the wild type *CO* gene into *co-8*, either by transformation or by crossing *co-8* with a transgenic line, will demonstrate whether all of the *co-8* phenotypes can be complemented by *CO*.

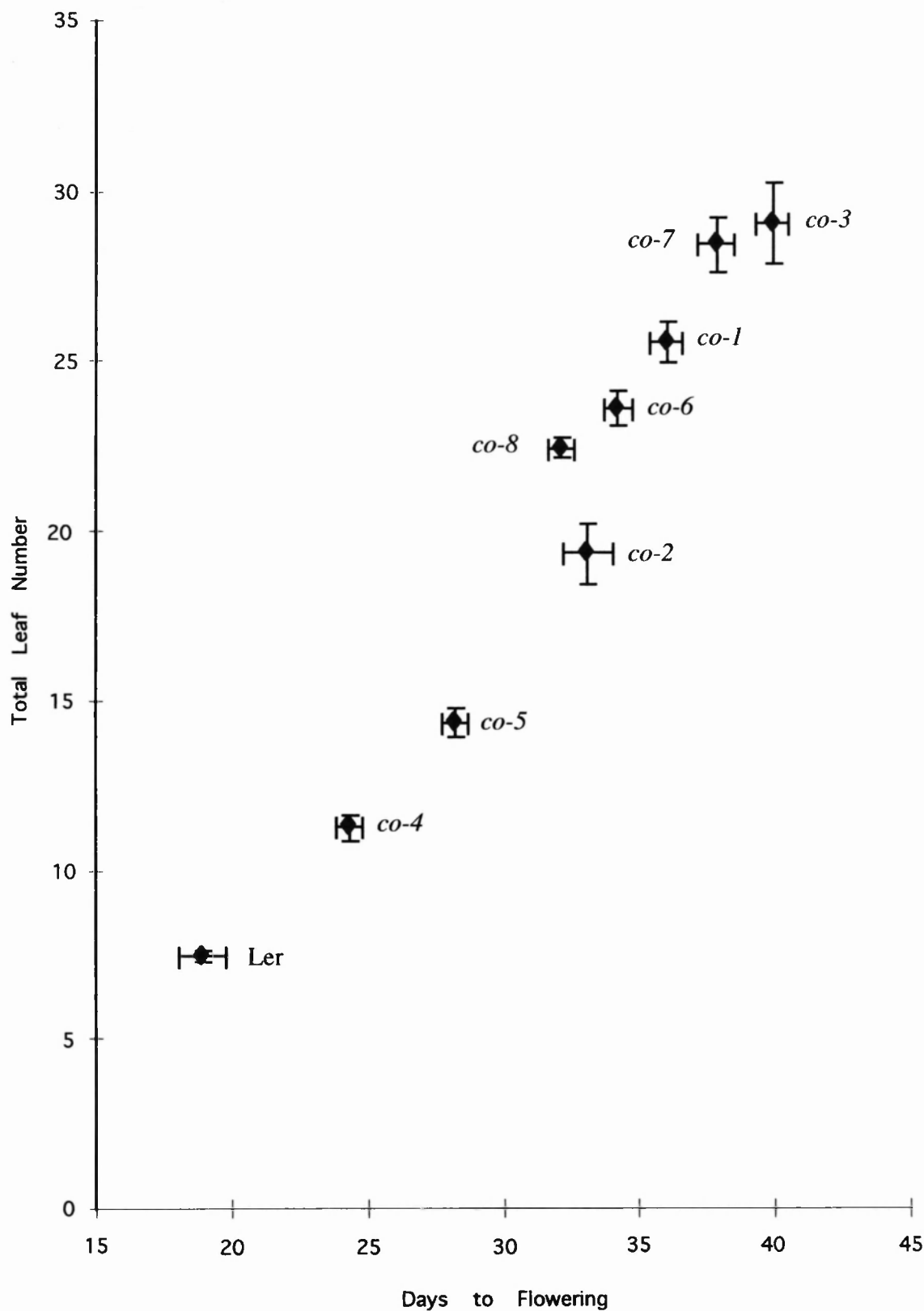


Figure 4.17 Days to flowering plotted against total leaf number for each *co* mutant and wild type grown in LDs. The *co-8* mutant has a higher leaf number than *co-2* but is induced to flower earlier. Standard error bars are in black.

4.3.2.3 *co-8* is semidominant

To determine whether, like the other *co* alleles, *co-8* is semidominant, F₁ seeds from a backcross to wild type Landsberg *erecta* were collected and sown in LD conditions. The flowering time of the F₁ plants was measured and compared with that of wild type and *co-8* homozygous plants (Figures 4.18 and 4.19). The *co-8* heterozygotes were intermediate in phenotype between *co-8* homozygotes and wild type plants with respect to flowering time and leaf number. They produced a flower bud after an average of 29.4 days and 16.2 leaves. Although *co-8* is one of the least severe mutants, the heterozygote flowers later and with more leaves than plants heterozygous for the other *co* alleles (Figure 4.19).

4.4 Discussion

A method for the isolation of deficiencies in the *Arabidopsis* genome was developed at Nottingham University (Vizir *et al.*, 1994). Gamma-irradiated wild type Landsberg *erecta* pollen was used to cross fertilize marker lines carrying a number of recessive mutations. This allowed the identification in the M₁ generation of new alleles at these loci on the irradiated chromosome. One of the marker lines carried the *co-1* allele. Late flowering M₁ plants from the cross of this line to irradiated pollen, were initially selected by Dr. Igor Vizir. Five of the twenty six M₂ populations analyzed did not segregate early flowering plants in either the M₂ or the M₃ generation and were therefore assumed to carry new *co* alleles. Two of these were completely sterile so work with them was discontinued. Previously, 45% of plants carrying mutations induced in a similar experiment were sterile and could not be maintained (Vizir *et al.*, 1994). The remaining three late flowering lines were screened by Southern analysis and a line renamed *co-8* was identified which carried a deletion at the *CO* locus. Southern analysis of the other two lines did not reveal any rearrangement. Possibly the γ -irradiation-induced alleles had been lost from these lines, perhaps because the deficiencies were large and not transmitted through the gametes.



Figure 4.18 The *co-8* mutant (right) compared to a *co-8* heterozygote (centre) and wild type *Landsberg erecta* (left). All plants were grown in LDs and are 28 days old.

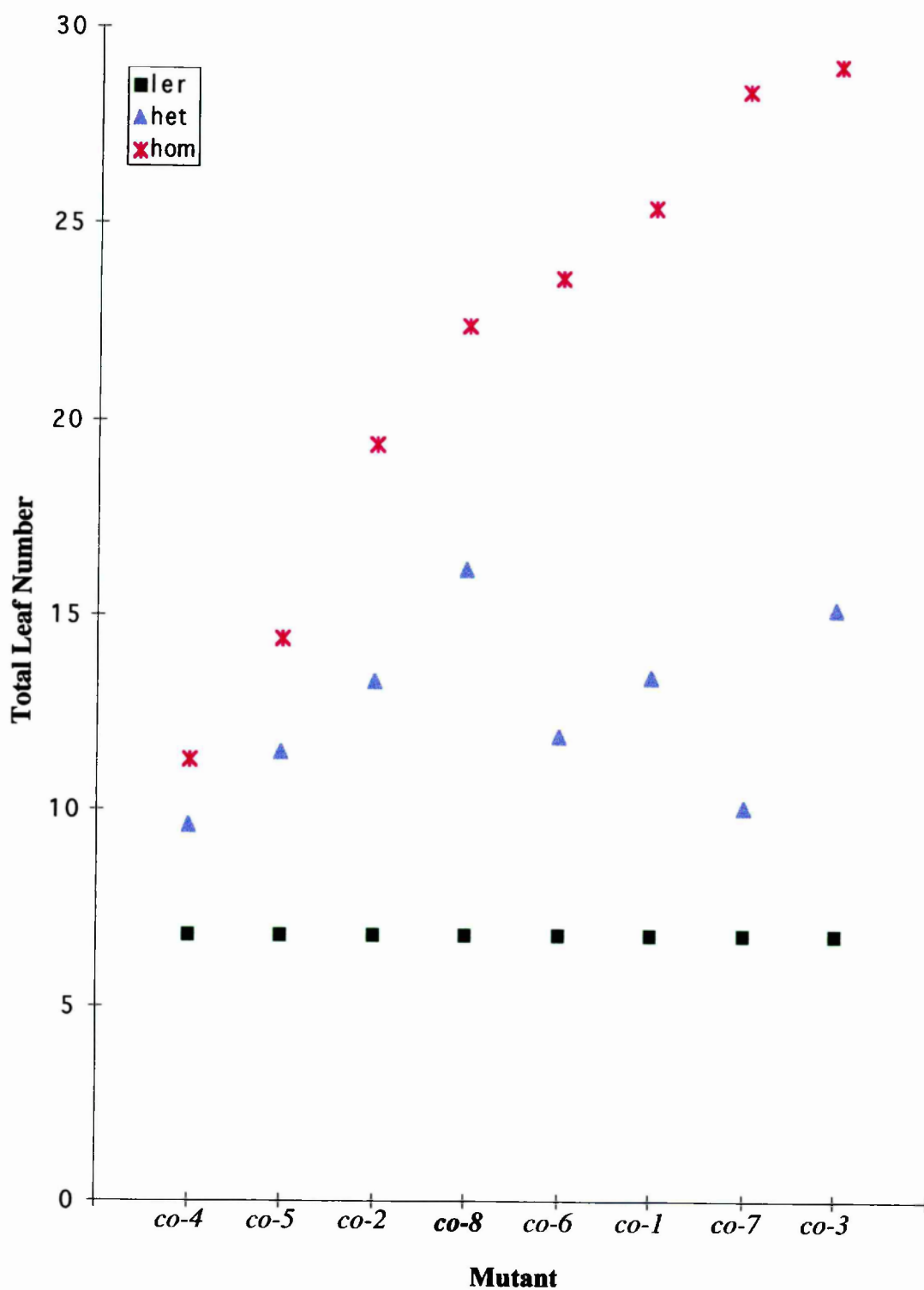


Figure 4.19 Summary of total leaf numbers of the *co* mutants grown in LDs, compared to the heterozygotes and wild type Landsberg *erecta*. *co-8* has been added to the series. The error bars are not included as they are smaller than the symbols.

Detailed molecular analysis of *co-8* revealed that as well as a deletion of 1.3 kb which removed the *CO* promoter and the coding sequence for the zinc finger domain, there was a second deletion at map position 115.2 cM on the lower arm of chromosome 5 and the intervening 82 cM, corresponding to a physical distance of 16.4 Mb was inverted. A transcript including the remaining 3' end of the *CO* gene was expressed in *co-8*. However the presence of an in-frame stop codon directly proximal to the portion of the *CO* ORF retained in *co-8* suggested that it was unlikely that a novel protein encoded by the retained portion of the *CO* ORF and part of the inverted DNA was made. Whether the remaining 3' end of the coding sequence was translated and still retained CO activity was not known, but the presence of a truncated CO protein in *co-8* could be tested for with an antibody.

A similar mutagenesis screen was used to find new *clavata1* alleles (Clark *et al.*, 1993). X-irradiated Columbia wild type pollen was used in a cross with *clv1-1 pi-1* double mutants in Landsberg *erecta*, and phenotypically *clv1* plants selected in the M₁ generation. Five *clv1* plants were identified, self-fertilized and analyzed in the next generation by Southern hybridization using RFLPs which mapped near to *clv1*. Three of the new alleles were homozygous for Landsberg *erecta* RFLPs indicating that the X-ray induced Columbia allele had been lost, possibly as a result of a large deletion that was unable to transmit through the gametes. One of the remaining new *clv1* alleles was sequenced and found to contain an extra A within the coding sequence which resulted in a frameshift mutation (Clark *et al.*, 1997).

Gamma-irradiation was used in other mutagenesis screens in *Arabidopsis* to isolate alleles at a particular locus of interest. Mutations in the nitrate reductase structural gene, *NIA2*, result in a chlorate resistant phenotype (Wilkinson and Crawford, 1991). New chlorate resistant mutants were generated by γ -irradiation of *Arabidopsis* seeds and mutants

identified in the M_2 generation. Three mutants contained deletions of more than 5 kb at the *NIA2* locus and these completely removed the gene.

The rearrangement in *co-8* is reminiscent of the rearrangements induced by ionizing radiation in two *Arabidopsis* genes which encode enzymes in the flavanoid biosynthesis pathway. Like *co-8*, the mutations not only affect the loci of interest but sites some distance away on the same chromosome. Mutations affecting the flavonoid biosynthetic pathway have been extensively studied in a number of plant species including maize, petunia and *Antirrhinum* (Stafford, 1990). They are non-lethal and have easily scorable phenotypes such as flower and seed colour. Eleven loci affecting this pathway have been isolated from *Arabidopsis* (Dellaert, 1980; Koornneef, 1990). They have a reduced amount or a complete absence of pigments in their seed coat and hence are named *transparent testa* or *tt*. The genes corresponding to two of the *tt* mutants, *tt3* and *tt5*, were cloned and the nature of the mutations at these loci determined (Shirley *et al.*, 1992). A *tt5* allele (40.443), generated by fast-neutron irradiation consisted of a 1.5 kb inversion within the *CHALCONE FLAVANONE ISOMERASE (CHI)* gene. A 272 bp fragment derived from a location 38 cM proximal to *tt5* (a physical distance of approximately 7600 kb) was translocated to one end of the inversion. A *tt3* allele (M218) was generated by X-irradiation and compromised three genes. Molecular characterization identified two deletions and an inversion, similar to the rearrangement in *co-8*. At the *tt3* end of the inversion a 7.5 kb deletion completely removed the *DIHYDROFLAVANOL 4-REDUCTASE (DFR)* gene and an adjacent gene. A second deletion of 52 bp, 2.8 cM distal to *DFR* lay in the middle of a third gene. The intervening 2.8 cM was inverted, a physical distance of approximately 560 kb.

The rearrangements in *co-8*, *tt3*-M218 and *tt5*-40.443 all involved four double strand breaks on the same chromosome and the subsequent aberrant rejoining of the ends to produce either two deletions and an inversion or a translocation and an inversion. The mutations in

tt5-40.443, which was generated by fast neutrons, were more restricted and less deleterious. Only one gene was affected and no deletion occurred. The *co-8* and *tt3-M218* alleles, both of which contained deletions were produced by γ -irradiation and X-irradiation, respectively. These two types of radiation produced sparse ionizations compared to the dense ionizations of fast neutrons (Hawkins, 1979). This was proposed to account for the differences in mutant spectra (different types of mutations) observed using different radiation sources (Dellaert, 1980).

The fact that the two sites affected in these three alleles were on the same chromosome perhaps reflects the position in the nucleus of the broken ends when the DNA repair mechanisms were operating. Although the region affected by the mutations in *co-8* was extensive, molecular analysis showed that the two deletions were small, and may only compromise *CO*. This might explain why *co-8* was recovered, while the other candidate *co* alleles, which might have carried larger deficiencies, were lost due to lack of transmissibility.

In the segregating M_2 generation, a distortion in favour of the recessive *lutescens* phenotype, carried on the non-irradiated parental chromosome, was observed in the *co-8* line. The ratio of *LU:lu* was 1:1.6 instead of the normal Mendelian segregation of a recessive phenotype of 3:1 (61.5% phenotypically *lutescens* plants instead of 25%). This suggested that a deficiency was present on the irradiated chromosome linked to the *LUTESCENS* locus (Vizir *et al.*, 1994) *LU* is located approximately 5.3 cM distal to *CO* (Putterill *et al.*, 1995). Semisterility of the M_1 heterozygous plant was also observed. This may have been due to the loss of the irradiated parental chromosome from some gametes. The large pericentric inversion in *co-8* may also reduce the viability of gametes. Recombination within this region will sometimes result in duplication and deletion of parts of the chromosome (Redei, 1992) leading to inviability of the gametes carrying these rearrangements. To confirm genetically that *co-8* contains an inversion a line homozygous

for *co-8* is currently being crossed to a line carrying a number of phenotypic markers on chromosome 5 (Igor Vizir, unpublished results). The presence of a pericentric inversion should repress recombination between these markers (Rédei, 1992). Crossovers between markers closely linked to the inversion breakpoints will be disrupted while crossovers within the inverted segment will produce inviable gametes due to duplication and deficiency.

The region 82 cM from the *CO* locus that was deleted in *co-8* was isolated and sequenced. It is not known whether it disrupts a second gene. If it does, this mutation, and consequently any phenotypic effect it has, cannot be separated from the *co-8* phenotype by backcrossing. Complementation by crossing *co-8* with a plant carrying the wild type *CO* gene on a T-DNA would distinguish between phenotypes caused by the mutation at the *CO* locus, and additional phenotypes caused by the *co-8* rearrangement.

The *co-8* allele has already proved useful during the analysis of *CO* expression by *in situ* hybridization. Primers which would specifically amplify the region of the *CO* transcript deleted from *co-8* were designed, and the resulting PCR product cloned and used to make an RNA probe. This was used against sectioned *co-8* meristems, as a negative control for expression, and against wild type meristems to monitor *CO* expression (R. Simon, unpublished results).

Primarily, the analysis of *co-8* has demonstrated that in the probable absence of any *CO* activity, *Arabidopsis* plants can still flower. In a *co* null allele, other flowering pathways, such as the autonomous flowering pathway which promotes flowering independently of environmental stimuli, may be able to compensate for the lack of *CO*. No mutants in *Arabidopsis* have been isolated that completely abolish flowering, which suggests that several pathways operate in parallel to promote flowering and all are partially redundant (Koornneef *et al.*, 1998; Chapter 1).

The *co-8* heterozygote is both later with respect to days to flowering and leaf number than the heterozygotes of all the other alleles, although this has not yet been verified in the F_2 generation. If *co-8* retains no CO function, then the earlier flowering of heterozygotes of the other *co* mutants could be due to the other mutant proteins retaining some CO function. However, the *co-8* homozygote has an intermediate phenotype, flowering earlier than five of the other *co* homozygotes. That *co-8* is not the most severe mutant might suggest that the proteins produced by the other *co* mutant alleles may antagonize the function of another gene whose ability to promote flowering is observed in *co* homozygotes. For example *co-7* may have a weak heterozygous phenotype because the mutant allele retains activity and a strong homozygous phenotype because the mutant protein strongly antagonizes the function of a second gene whose activity is most important in *co* homozygotes.

Alternatively, *co-8* may still retain some CO activity if the remaining C-terminus of the protein is translated. This may be exerting a dominant effect in the presence of a wild type copy of CO in the heterozygote. The dominant effect could be due to the truncated *co-8* protein titrating wild type CO into a non-functional dimer, or a completely new function of the truncated protein. Once an antibody specific for the CO protein is available, this could be used to determine whether any truncated CO protein is present in *co-8*. As mentioned in Chapter 3, experiments are currently underway to determine the function of the important domains of CO. Once the wild type functions of these domains have been analyzed it may be possible to examine how the functions are compromised in the mutant alleles. This may indicate whether the mutant alleles might be loss of function alleles or dominant gain of function alleles.

To examine the nature of the *co-8* mutation a reverse-complementation approach could be taken. A genomic fragment containing the *co-8* allele could be introduced by transformation into wild type plants. This would necessitate re-screening the library made from *co-8* DNA to identify a cosmid containing the remainder of the *CO* ORF and sufficient upstream

sequence to drive its expression. If *co-8* is a null allele, then introduction of this allele into a background containing two copies of the wild type *CO* gene should not affect flowering time. If, on the other hand, *co-8* is a dominant gain of function allele, then it might be expected to interfere with wild type CO function and delay flowering.

The observation that either *co-8* itself or some of the other alleles could be exerting a dominant effect on flowering, and the analysis in the previous chapter of the discrete functional domains of the CO protein suggested that expression of modified CO proteins might repress flowering. To further gain an insight into the role of *CO*, the data obtained from the analysis of the *co* alleles was used to design modified CO proteins which, when introduced by transformation into wild type *Arabidopsis thaliana*, might be expected to delay flowering.

5 Overexpression of modified CONSTANS proteins

5.1 Introduction

The CONSTANS protein has a number of structural features characteristic of zinc binding transcription factors. Although these functions have not yet been demonstrated, the CO protein was proposed to have domains required for DNA binding, transcriptional activation, nuclear localization and protein-protein interaction. In other proteins, the arrangement of the different functions of the protein into discrete domains has made it possible to design inhibitory proteins which retain one or more of the protein's functions but are compromised in one of the others. These "dominant negative" proteins may be used to identify the function of a gene by inhibiting the wild type gene product (reviewed in Herskowitz, 1987). An example in *Arabidopsis* was described by McNellis *et al.* (1996) for the CONSTITUTIVE PHOTOMORPHOGENIC1 (COP1) protein. This protein contains a RING finger zinc binding domain, a coiled-coil domain and a WD-40 repeat domain (von Arnim and Deng, 1993). Ectopic expression in transgenic *Arabidopsis* of a derivative encoding only the N-terminal RING finger zinc-binding and coiled-coil domains caused a dominant negative phenotype resembling the loss-of-function *cop1* mutants (McNellis *et al.*, 1996). Overexpression of the wild type CO protein in transgenic *Arabidopsis* plants dramatically accelerates the transition to flowering in both long and short days (Simon *et al.*, 1996; Igeño *et al.*, unpublished results). The creation of an inhibitory CO protein which interfered with normal wild type CO function might delay flowering of transgenic plants.

5.2 Description of the modified CONSTANS proteins

When the *co* alleles were sequenced it was found that the two most severe alleles, *co-3* and *co-7*, contained missense mutations that each affected one of the two very highly conserved domains of the CO protein (Chapter 3, Figure 3.6). If each of these mutations destroys the function of one domain, but leaves the remainder of the protein intact, their overexpression might cause a dominant negative effect. Both alleles are more severe than the putative null allele, *co-8*, (Chapter 4) so it is possible that the phenotype of these late alleles is caused by dominant inhibition of flowering rather than a simple loss of function.

The *co-3* mutation is within the zinc finger region. If this protein cannot bind DNA, but can still interact with other proteins necessary for CO function, possibly through the conserved basic domain at the C-terminus (Chapter 3, Figure 3.11), then excess *co-3* protein might sequester all the available CO-interacting protein into a non-functional complex. The endogenous CO protein would therefore not be able to function and transgenic plants overexpressing *co-3* would flower late.

The *co-7* allele, on the other hand, encodes a normal DNA binding domain, but carries a point mutation affecting the conserved C-terminal basic domain. This domain may have a role in protein-protein interactions. If this mutant protein was overexpressed in a wild type plant the excess *co-7* protein may occupy the available DNA binding sites but fail to interact with other factors important for CO function. This might therefore inhibit the function of the wild type protein and cause late flowering.

Similar effects to those discussed above for *co-3* and *co-7* may be achieved by completely removing domains of the protein. For example, removing the C-terminus of the protein, leaving only the N-terminus that includes the two zinc fingers and the conserved tail thought to be important for DNA binding might produce a truncated protein that binds DNA but does not interact with other proteins. This would be particularly useful if *co-7*, which is slightly less severe than *co-3*, retains some ability to promote flowering.

The three constructs overexpressing *co-3*, *co-7* and the truncated protein were designed so that the strong Cauliflower Mosaic Virus 35S promoter (35S) was fused to the *co-7* or *co-3* ORFs, or to the coding region for the zinc fingers (ZF) only (Figure 5.1). Their construction is described below.

5.3 Plasmid construction

The *CO* gene was amplified by PCR in duplicate reactions from all of the alleles. A fragment derived from each amplification was blunt ended and cloned into the *EcoR* V site of pBluescript (SK+) and sequenced (Chapter 3). Clones which contained no PCR errors introduced by *Taq* polymerase were identified for *co-3* and *co-7* and used in the plasmid constructions. The construct containing the CaMV 35S promoter in front of the wild type *CO* gene in the binary vector SLJ1711 (Jones *et al.*, 1992) was kindly provided by Maribel Igeño.

5.3.1 35S:*co-7*

To clone the *co-7* allele, primers *co41* and *co42* were used to amplify a 1.95 kb genomic fragment (Chapter 3) containing both exons, the single intron, 250 bp of 5' sequence and 350 bp of 3' sequence including the *CO* polyadenylation signal. After sequencing clones derived from separate PCR reactions a clone which did not contain any PCR errors - *p479* - was identified. This was cloned in the correct orientation in the *EcoR* V site such that it could be used to replace the wild type *CO* genomic region in the existing binary vector-based plasmid *p505* (Igeño *et al.*, unpublished results; Figure 5.1). A *Hind* III - *Bam*HI digestion of *p505* excised a 1.79 kb fragment which could be replaced with the equivalent *co-7* genomic fragment from *p479*. However, an extra unexpected *Hind* III site was found in the binary vector SLJ1711, so an alternative strategy was used. The 35S promoter, originally derived from construct *pJIT62* (Guerineau *et al.*, 1991), was removed from *p505*

using *Cla* I and *Hind* III. The p479 *co-7* clone was digested using *Cla* I, which cuts in the polylinker at the 5' end of the gene, and *Hind* III, which cuts 70bp upstream of the *CO* ATG, a site which has been used to create several transcriptional fusions between *CO* and foreign promoters. The gel-purified 35S promoter fragment was then ligated to the gel purified construct *Cla* I - *Hind* III fragment of p479 to make p35S:*co-7* (Figure 5.1). To move the entire 35S:*co-7* fusion into the SLJ1711 binary vector, both p35S:*co-7* and pSLJ1711 were digested with *Cla* I and *Bam* HI and the 35S:*co-7* containing fragment ligated to the binary vector.

5.3.2 35S:*co-3*

The cloning strategy to make the 35S:*co-3* construct was essentially the same as the one described for the 35S:*co-7* construct, except that by chance both *co-3* PCR fragments were only cloned in pBluescript in the orientation that was inappropriate for the cloning strategy described for *co-7*. Therefore the orientation of the *co-3* genomic region was changed before cloning into the binary vector. This was done by first cloning the 35S promoter *Cla* I - *Hind* III fragment into *Cla* I - *Hind* III digested pBluescript (SK+) to make p35S. The *co-3* genomic clone, p470, was digested with *Hind* III, which cuts in the polylinker at the 3' end of the gene as well as 70bp upstream of the ATG, as described previously, yielding a 1.7 kb fragment. This was gel-purified and ligated to *Hind* III digested p35S. The orientation of the *co-3* *Hind* III fragment was determined by restriction mapping. The resulting p35S:*co-3* clone (Figure 5.1) was digested with *Cla* I and *Bam* HI and the 35S:*co-3* fragment ligated to pSLJ1711 as previously described.

5.3.3 35S:ZF (Zinc Fingers)

To make a construct containing only the sequence coding for the putative DNA binding domain of *CO*, a PCR primer was designed to *CO* sequence 3' of the zinc fingers. When used with an existing PCR primer that annealed 5' of the *CO* ATG, the 5' end of the *CO* gene, including sequence encoding both zinc fingers, as well as DNA encoding the highly conserved region to the carboxyl side of the second finger and an additional 17 amino acids was amplified. The primer 3' of the zinc finger region also had an in-frame stop codon engineered near its 5' end, and three additional bases to ensure that the stop codon was not destroyed during blunt-end cloning (Figure 5.2). This primer, co65 (Chapter 7, Materials and Methods) was used with primer co41 to amplify a 630 bp fragment in duplicate PCR reactions using DNA extracted from wild type Landsberg *erecta* plants. Fragments derived from duplicate reactions were blunt-end cloned into the *EcoR* V site of pBluescript and their orientation determined by restriction mapping. Clones containing the insert in the correct orientation, such that the *Cla* I site in the polylinker was at the 5' end of the gene, as before, were sequenced to check for PCR errors, and to verify that the introduced stop codon was in frame with the rest of the *CO* sequence. The *Cla* I - *Hind* III fragment containing the 35S promoter was inserted into the *Hind* III site upstream of the ATG, as previously described. As the truncated version of the *CO* gene in this construct did not contain any of its 3' untranslated sequence, the 35S termination and polyadenylation sequences, from vector pJIT62, were isolated as a 720bp *Sma* I - *Eco* RV fragment. This was inserted at the 3' end of the truncated gene into the *Sma* I site and its orientation determined by restriction mapping. The resulting construct, p35S:ZF (Figure 5.1), was digested with *Cla* I and *Bam* HI and the 35S:ZF:polyA fragment ligated to SLJ1711 as previously described.

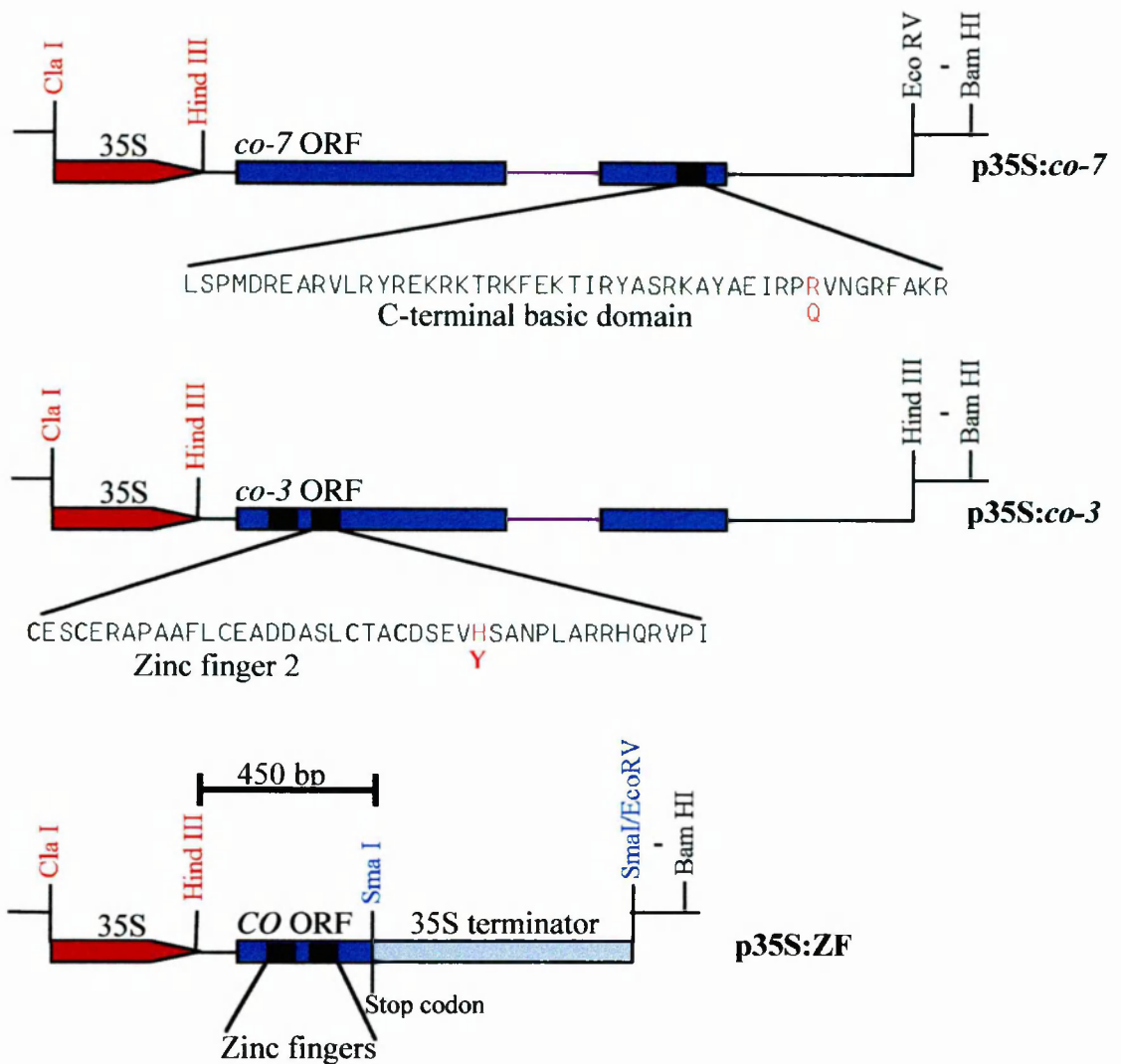


Figure 5.1a Restriction maps of the constructs designed to express modified CO proteins. All constructs were created in pBluescript before the completed constructs were moved into the binary vector SLJ1711 (Jones *et al.*, 1992).

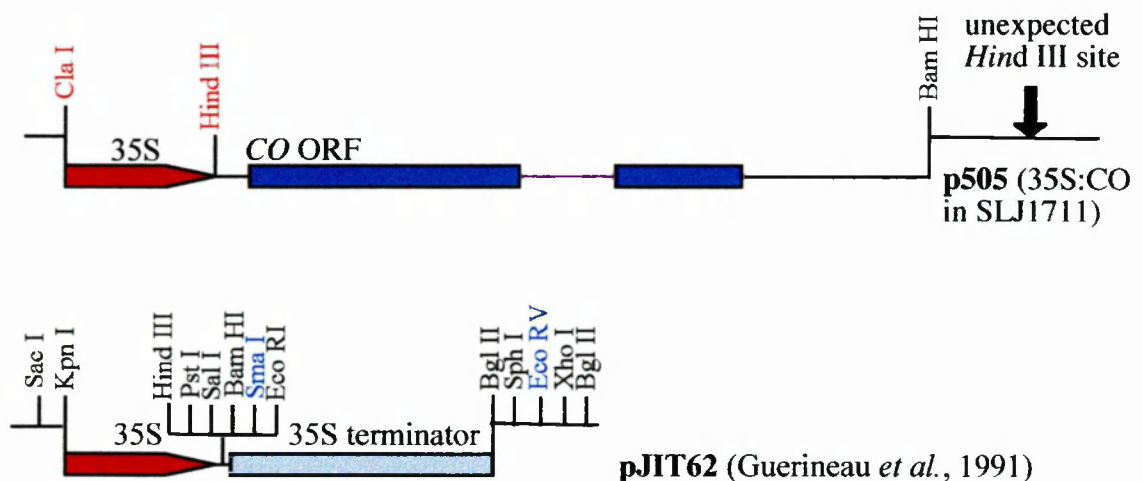


Figure 5.1b Restriction maps of the two constructs that provided the 35S promoter (p505) and 35S termination and polyadenylation signal (pJIT62).

c041 ggtcccaacgaagaagtgc

gaggagatgtttacactttacactttacactttacactttacacataggccttcccaaaa
gctcaactagctgcaagaggatccaataacatgtaagagccactaacgctgtgccacgtg
taggcactcaggattcgatcttccctctacttattctctcacaccagatatagcttta
ttagcccttctttcagataccagctcccacaccatcaaacttactacatctgagttatt

atgttgaaacaagagagtaacgacataggtagtggagagaacaacagggcacgacctgt
-----+-----+-----+-----+-----+-----+-----+
tacaactttgttctctcattgctgtatccatcacctctcttgttgtcccgtgctgggaca

M L K Q E S N D I G S G E N N R A R P C

gacacatgccgggtcaaacgcctgcaccgtgtattgccatgcagattctgcctacttgtgc
-----+-----+-----+-----+-----+-----+-----+
ctgtgtacggccagtttgcgacgtggcacataacggtacgtctaagacggatgaacacg

D T C R S N A C T V Y C H A D S A Y L C

atgagctgtgatgctcaagttcactctgccaatcgcggttgcctcccgccataaacgtgtc
-----+-----+-----+-----+-----+-----+-----+
tactcgacactacgagttcaagtgagacggttagcgcaacgaagggcggtatttgcacag

M S C D A Q V H S A N R V A S R H K R V

cgggtctgcgagtcattgtgagcgtgctccggctgcttttttgtgtgaggcagatgatgcc
-----+-----+-----+-----+-----+-----+-----+
gccagacgctcagtaactcgcacgaggccgacgaaaaaacacactccgtctactacgg

R V C E S C E R A P A A F L C E A D D A

tctctatgcacagcctgtgattcagaggttcattctgcaaaccacttgctagacgccaat
-----+-----+-----+-----+-----+-----+-----+
agagatacgtgtcggacactaagtctccaagtaagacggttgggtgaacgatctgcggtta

S L C T A C D S E V H S A N P L A R R H

cagcgagttccaattctaccaatttctggaaactctttcagctccatgaccactactcac
-----+-----+-----+-----+-----+-----+-----+
gtcgtctcaaggttaagatggttaaagacctttgagaaagtcgaggtact **ggtgatgagtg**

Q R V P I L P I S G N S F S S M T T H

caccaaagcgag**tga**cgc
-----+-----
gtggtttcgctcactgcg
← c065
H Q S E * R -

Figure 5.2 The sequence of the DNA fragment amplified to create the 35S:ZF construct. The primers used are shown in bold. Primer c065 was designed to introduce an in-frame stop codon (TGA). The ORF sequence is shown in both strands, with the amino acid sequence underneath. The putative DNA binding domain is underlined. The *Hind*III site used to introduce the 35S promoter in front of the ORF is in a box.

5.4 Introduction of the constructs into *Arabidopsis thaliana*

The three constructs were transferred to *Agrobacterium tumefaciens* strain C58C1 pGV2260 (Zambryski *et al.*, 1983; Deblaere *et al.*, 1985) by tri-parental mating. The integrity of the plasmids in the *Agrobacterium* transconjugants was verified by restriction digestion and Southern analysis, and they were introduced into wild type *Arabidopsis thaliana* ecotype Landsberg *erecta* by root transformation or vacuum infiltration.

Transformations with constructs 35S:co-7, 35S:co-3 and 35S:ZF gave 38, 18 and 19 independent transformants, respectively. The T₂ seeds from the primary transformants (T₁) were sown on GM containing kanamycin and the ratio of kanamycin resistant:sensitive seedlings recorded to confirm that the T₁ plants were transformed and to ascertain the number of loci at which the T-DNA had inserted (Table 5.1).

Transformations using constructs 35S:co-3, 35S:ZF, and the *Brassica napus* Bn101 construct, described in Chapter 6, were performed simultaneously, and yielded a surprisingly high number of families containing no kanamycin resistant plants. These were presumably derived from T₁ plants that were not transformed and this problem was not encountered in any other experiment. It was assumed to reflect the age or condition of the root material used for this experiment or a problem with a particular batch of hormones or antibiotics. There was not sufficient time to repeat these transformations.

For 35S:co-3, 4/18 T₁ plants and for 35S:ZF, 5/19 T₁ plants gave kanamycin resistant progeny. In contrast only one out of 38 transformants for the 35S:co-7 construct did not produce kanamycin resistant progeny. The kanamycin resistant T₂ plants were transferred to the greenhouse once they had produced a few leaves, self fertilized and the T₃ seeds collected from each T₂ plant to determine its genotype with respect to the T-DNA. The T₃ families were also used for flowering time experiments.

<u>Construct</u>	<u>Transformant</u>	<u>Km^r:Kms segregation in T₂</u>	<u>T3 family</u>	<u>Km^r:Kms segregation in T₃</u>	<u>Comments</u>
35S:<i>co-3</i>	Tn101	3.0:1	Tn101-2	5.2:1	
			Tn101-4	1:0	Hom
	Tn104	2.6:1	Tn104-4	2.8:1	
			Tn104-5	1:0	Hom
	Tn112	3.0:1	Tn112-3	3.0:1	
			Tn112-4	1:0	Hom
	Tn114	3.0:1	Tn114-2	4.0:1	
			Tn114-4	1:0	Hom
35S:<i>co-7</i>	Tn33a	2.6:1	Tn33a-3	2.9:1	
			Tn33a-1	1:0	Hom
	Tn43	3.2:1	Tn43-10	3.8:1	
			Tn43-1	1:0	Hom
	Tn75a	3.5:1	Tn75a-5	3.3:1	
			Tn75a-1	1:0	Hom
	TnG209	3.3:1	TnG209-11	4.5:1	
			TnG209-8	1:0	Hom
	TnG210	1.9:1	TnG210-18	2.8:1	
			TnG210-7	1:0	Hom
35S:ZF	Tn204	1.3:1	Tn204-3	11:1	
			Tn204-2	1:0	Hom
	Tn206	34:1	Tn206-1	47:1	
			Tn206-2	1:0	Hom
	Tn211	1.0:1	Tn211-2	2.9:1	
			Tn211-4	1:0	Hom
	Tn213	50:1	Tn213-4	106:1	
			Tn213-2	1:0	Hom
	Tn215	24:1	Tn215-2	3.0:1	
			Tn215-1	1:0	Hom

Table 5.1 The kanamycin segregation analysis of the transformants in the T₂ and T₃ generations. Only two T₃ families are shown from the five analyzed for each transformant: a representative family heterozygous for the T-DNA/s and the homozygous family chosen for further flowering time analysis.

5.5 Identification of homozygous families carrying the T-DNA at a single locus

Approximately 100-150 T₃ seeds from five T₂ families for each transformant were sown on media containing 50mg/L kanamycin, and after 1-2 weeks, the ratio of resistant:sensitive seedlings was recorded (Table 5.1). Where possible, transformants thought to carry the T-DNA at a single locus were identified because they produced T₂ and two thirds of T₃ families segregating kanamycin resistance:sensitivity in a 3:1 ratio. From these transformants, T₃ families showing 100% kanamycin resistance, and therefore homozygous for the T-DNA, were identified and analyzed further to ascertain whether the construct they contained delayed the onset of flowering.

5.6 Phenotypic analysis of the transformants under long and short days

Apart from flowering time which is discussed in detail below, a phenotype seen frequently among T₃ families was an increase in the number of petals of the first or second fully opened flower on the primary inflorescence, as described for the *co* mutants (Chapter 2, Figure 2.8). The phenotype was observed among plants carrying all three constructs but not in all independent transformants. For the 35S:*co*-7, 35S:*co*-3 and 35S:ZF constructs, grown in long days, 1/5, 2/4 and 5/5 transformants, respectively, showed the phenotype in 5 - 20% of individuals. The phenotype was not observed in any of the long day-grown Landsberg *erecta* control plants.

To analyze the flowering time phenotypes, 20 seeds from homozygous T₃ families of five of the thirty seven 35S:*co*-7 transformants, the four 35S:*co*-3 and the five 35S:ZF transformants were grown in LD conditions beside Landsberg *erecta* plants. Flowering time was scored as the number of days from sowing until the first appearance of buds in the centre of the rosette, and total leaf number, counted after the plants had bolted.

Although it was evident that none of the constructs had a dramatic effect on flowering time, a small number of transformants did appear to show either slightly delayed or accelerated

flowering (Table 5.2). To establish whether these observations were statistically significant the flowering time and leaf number data were subjected to one-way analysis of Variance, (ANOVA, Minitab Statistical Software PC version release 8, Minitab Inc. 1991) to ask if the flowering time/leaf number of plants carrying each construct was significantly different from the untransformed control.

For both flowering time (days to induction) and leaf number of the individual transformants compared to untransformed *Landsberg erecta*, the analysis of variance indicated that highly significant differences were present ($P < 0.001$). For a pairwise comparison of each transformant, or each construct, to *Landsberg erecta*, the significance level was therefore set at 0.001 (0.1%) to look for highly significant differences. The output from ANOVA is detailed in the Appendix.

5.6.1 35S:co-7

When the mean flowering times of the five independent 35S:co-7 transformant populations were subjected to a one way analysis of variance only one was significantly earlier than *Landsberg erecta* at the 0.1% level. Transformant G209 was induced in LDs to flower after 14.2 days from germination while *Landsberg erecta* was induced after 20.1 days (Table 5.2). When the mean flowering times for all transformants carrying this construct were pooled there was no significant difference at the 0.1% level between this value and the mean flowering time for the *Landsberg erecta* population. However, similar comparisons between 35S:co-7 and 35S:co-3 demonstrated that 35S:co-7 transformants had significantly fewer leaves than 35S:co-3 transformants, at the 0.1% level, which confirmed the initial observations that 35S:co-7 slightly accelerates flowering and 35S:co-3 slightly delays flowering.

The observation that 35S:co-7 accelerates flowering was reconfirmed when these transformants were grown in short day conditions where *CO* is limiting. In short day

conditions the *CO* transcript is less abundant than in long days (Putterill *et al.*, 1995). Early flowering under short days can be caused by overexpressing *CO* behind the strong 35S promoter (Igeño *et al.*, unpublished results) or by introducing multiple copies of the wild type *CO* gene (Putterill *et al.*, 1995). To establish whether the 35S:*co-7* construct was able to accelerate flowering the transformants were grown in SDs along with Landsberg *erecta* controls, and flowering time and leaf number recorded as before. All transformants flowered earlier and with fewer leaves than the control plants (Table 5.2; Figure 5.3). Three out of five transformants flowered significantly earlier than wild type and all five transformants flowered with significantly fewer leaves than wild type, at the 0.1% level (Appendix). The relative flowering times of the 35S:*co-7* transformants were the same under LDs and SDs when they were compared to each other or to wild type. For example, G209, the earliest in long days, was also the earliest in short days, being induced to flower after 34.3 days with 16.5 leaves. The Landsberg *erecta* control flowered after 60.3 days with 48.9 leaves (Table 5.2; Figure 5.3).

The acceleration caused by the 35S:*co-7* construct in both long and short days is not as dramatic as the effect of expressing the wild type *CO* gene from the 35S promoter (Igeño *et al.*, unpublished results). 35S:*CO* transformants are induced to flower in long days after an average of 13.6 days from germination with 5 leaves and in short days after 15.7 days with 4.8 leaves. However there are other similarities between 35S:*co-7* and 35S:*CO*. For example, an increase in determinacy with premature termination of the shoot in a pistil and, in the more extreme 35S:*co-7* transformants, a reduction in height and flower number. The terminal pistil phenotype occurred most frequently among progeny of transformant G209, the earliest flowering family, but also among progeny of transformant G210, which was the latest flowering transformant carrying this construct.

5.6.2 35S:co-3

The transformants carrying the 35S:co-3 construct again showed variation in both their flowering times and total leaf number and some transformants were not significantly different to wild type. When the flowering time, measured in days from sowing to first visible buds, for the four transformants and wild type were subjected to a one way analysis of Variance (ANOVA), TN114 was significantly earlier than wild type, and TN101 was significantly later, at the 0.1% level. The remaining two were not significantly different to wild type. However, analysis of total leaf number at flowering indicated that no transformants flowered with fewer leaves, and again TN101 stood out as being slightly delayed in that it had significantly more leaves (9.8) than wild type (7.2) at the 0.1% level. TN114, which appeared to form floral buds after fewer days than wild type also had a slightly higher leaf number than wild type, although this was not significant at the 0.1% level. When the mean leaf numbers for all transformants carrying 35S:co-3 were pooled they were found to have significantly more leaves than wild type and the 35S:ZF and 35S:co-7 transformants at the 0.1% level.

5.6.3 35S:ZF

The flowering times and leaf numbers of the five 35S:ZF transformants were not significantly different from the wild type controls with one exception; Tn204 flowered with significantly more leaves than wild type, at the 0.1% level. However, individuals from all five 35S:ZF transformant populations showed the same flower phenotype seen in a number of 35S:co-7 and 35S:co-3 transformants. When grown in long days, between 5 and 20% of individuals from the five 35S:ZF transformant populations had more than the usual four petals in one or two of the first flowers on the primary inflorescence. If this is a phenotype caused by the transgenes and is not a consequence of regeneration through tissue culture

Transformant	Long Days		Short Days	
	Days to induction of flowering	Total leaf number	Days to induction of flowering	Total leaf number
35S:c o- 3				
Tn101-4	22.3±0.3	9.8±0.2	NT	
Tn104-5	18.4±0.3	7.6±0.2		
Tn112-4	18.9±0.3	7.3±0.2		
Tn114-4	17.7±0.2	8.1±0.2		
35S:c o- 7				
Tn33a-1	18.9±0.3	7.0±0.1	51.9±1.3	35.7±1.4
Tn43-1	19.9±0.1	6.4±0.2	53.5±0.9	36.9±1.1
Tn75a-1	21.3±0.2	7.9±0.2	57.5±0.7	39.4±0.7
TnG209-8	14.2±0.1	6.4±0.1	34.3±0.5	16.5±0.2
TnG210-7	21.3±0.2	7.9±0.3	56.9±1.3	42.2±0.7
35S:ZF				
Tn204-2	20.2±0.5	8.5±0.2	NT	
Tn206-2	20.4±0.4	8.1±0.3		
Tn211-4	19.8±0.6	7.6±0.2		
TN213-2	20.7±0.2	6.9±0.2		
Tn215-1	19.5±0.4	6.9±0.2		
Wild type <i>Ler</i>	20.1±0.5	7.2±0.3	60.3±1.2	48.9±0.8
<i>co-2 tt4</i>	30.9±2.0	14.4±1.0	48.8±0.8	36.7±1.1

Table 5.2 Flowering time and leaf number analysis of the transformants in LDs and, where examined, SDs, compared to untransformed wild type *Landsberg erecta* and the *co-2* mutant.



Figure 5.3 A wild type Landsberg *erecta* plant on the left, and a plant carrying the 35S:co-7 construct on the right (transformant TnG209). Both plants were grown in SDs and are 60 days old.

then the truncated CO protein must be expressed from the transgene and have an effect in developing flowers.

5.7 Analysis of the expression of the transgenes by Northern hybridization

To verify that the transgenes were intact and expressing the modified *CO* mRNAs, RNA was extracted from progeny of each independent transformant and analyzed by Northern hybridization using the *CO* cDNA as a probe. All transformants carrying 35S:*co-3* or 35S:*co-7* were found to strongly express a *CO* transcript. In the 35S:ZF transformants two transcripts were detected, both of which were much less abundant than the transcript detected in the 35S:*co-3* and 35S:*co-7* transformants. No transcript was detected in the Landsberg *erecta* untransformed control, which is expected as the endogenous *CO* transcript is present at a level undetectable by Northern hybridization. As expected, the transcript present in the 35S:*co-7* and 35S:*co-3* lines was approximately 1.3 kb. One of the two transcripts present in the 35S:ZF transformants was of the predicted size of 650 bp (450 bp of the *CO* ORF and approximately 200 bp of the 35S terminator up to the polyadenylation signal). The additional *CO*-hybridizing transcript in these transformants was approximately 1.5 kb.

5.8 Discussion

Naturally occurring dominant negative mutations were described by Muller (1932) as "antagonistic mutant genes, having an effect actually contrary to that of the gene from which they were derived". Well characterized examples include mutations affecting the multimeric *lac* and *trp* operon repressor proteins of *E. coli*, where the DNA binding domains are inactivated but the domains required for oligomerization are functional. In the presence of the wild type protein, mixed, non-functional aggregates form, which are unable to bind to DNA (Miller, 1978). Examples also exist in eukaryotes: in maize there is a

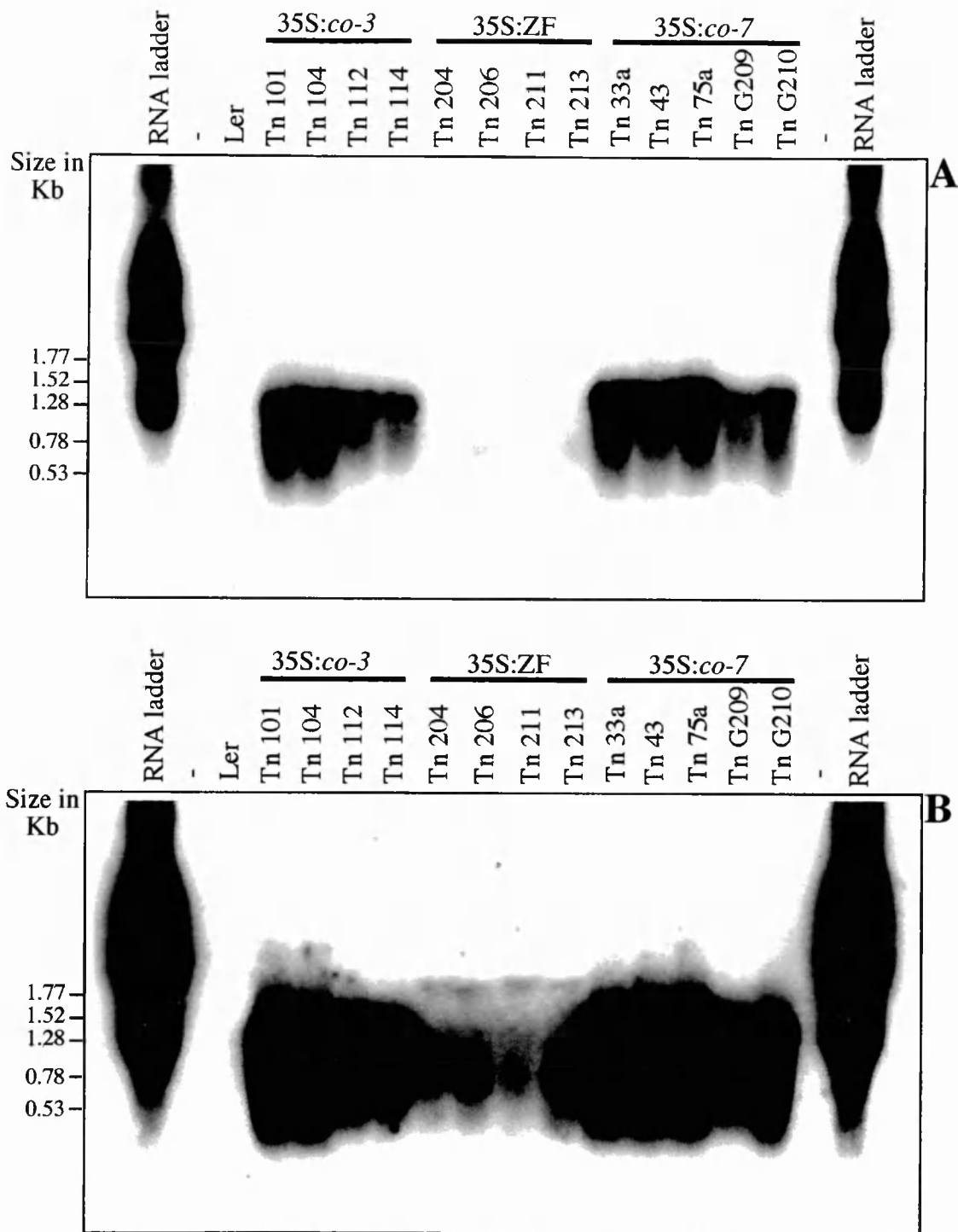


Figure 5.4 Northern blot of RNA extracted from the 35S:co-3, 35S:ZF and 35S:co-7 transformants. The blot was probed with the *CO* cDNA. Two different exposures are shown:

A) 8 hour exposure to X-ray film, showing the single 1.3 Kb transcript present in the 35S:co-3 and 35S:co-7 transformants. The two transcripts present in the 35S:ZF transformants are just visible.

B) Three day exposure to X-ray film showing the presence of at least two *CO*-hybridizing transcripts in the 35S:ZF transformants. No hybridizing transcript was detected in the Landsberg *erecta* control, as expected.

naturally occurring dominant negative mutation of the anthocyanin biosynthesis regulatory gene *Cl* (Paz-Arez *et al.*, 1990; Goff *et al.*, 1991). The *Cl-I* allele blocks pigmentation in the cells of the aleurone layer in the *Cl/Cl-I* heterozygote. By creating fusion proteins between the wild type *Cl* and the mutant *Cl-I* allele in a transient expression system, it was found that mutations in both the N-terminal DNA binding domain and the C-terminal acidic activation domain contribute to the dominant phenotype. The mutant protein may inhibit binding of the functional wild type protein to DNA by occupying the binding sites, or may interact directly with the wild type protein to form an inactive complex. The analysis of the yeast transcriptional machinery is well documented and the construction of artificial dominant negative mutations has helped in the functional dissection of many of the genes involved. For example, the separation of the DNA binding domain from the transcriptional activation domain of GCN4 (Hope and Struhl, 1986) and GAL4 (Keegan *et al.*, 1986) resulted in some of the derivatives acting as transcriptional repressors which could bind to DNA, preventing the binding of wild type proteins, but could not activate transcription.

In plants, one of the first proteins to be analyzed *in vitro* by the artificial construction of dominant negative derivatives was the maize OPAQUE2 bZIP transcriptional activator (Unger *et al.*, 1993). This protein regulates the genes encoding the 22-kD zein class of storage proteins in the maize endosperm. Derivatives of OPAQUE2, which had the basic DNA binding domain deleted formed heterodimers with wild type OPAQUE2 proteins when they were co-expressed in an expression system, and these heterodimers could not bind to the 22-kD zein promoter. Removal of the putative transactivation domain allowed the formation of both homo- and heterodimers with the wild type proteins, both of which could bind to the promoter but not activate transcription.

Subsequently, a number of other plant proteins that have a modular structure have been analyzed by the creation of dominant negative mutations, both *in vitro* and *in vivo*. These include Phytochrome A from rice (Emmler *et al.*, 1995), S₃ RNase from *Petunia* (McCubbin *et al.*, 1997), the bZIP transcription factor PG13 from potato (Rieping *et al.*, 1994), and several *Arabidopsis* genes including Cdc2a kinase (Hemerly *et al.*, 1995), Phytochrome A (Boylan *et al.*, 1994), Phytochrome B (Wagner *et al.*, 1996), the zinc-binding regulatory protein COP1 (McNellis *et al.*, 1996) and the MADS box regulatory protein AGAMOUS (Mizukami *et al.*, 1996).

The aim of the experiments described in this chapter was to study the *in vivo* function of individual domains of the CO protein by introducing dominant negative forms which may inhibit wild type CO function in transgenic plants. It was hoped that this would delay the onset of flowering, which, as well as providing information on the function of CO, could have major implications for the manipulation of similar genes in commercial crop plants. However, with the exception of small effects in one or two transformants, little or no significant alteration in flowering time phenotype was observed.

There are several possible explanations for why the transformants did not show clear phenotypes. The first is purely technical and concerns the 35S:ZF construct. Although the construct was thoroughly checked by restriction mapping when transferred into *Agrobacterium*, it is possible that it became rearranged subsequently in the transformants. Northern analysis showed that at least two CO-hybridizing transcripts were present in these transformants, one of the predicted size (650 bp, consisting of 450bp of CO sequence and approximately 200bp of additional 3' sequence derived from the CaMV polyadenylation sequence (Guerineau *et al.*, 1991)) and an additional one of approximately 1.5 kb. Both transcripts in the 35S:ZF transformants were also considerably less abundant than those

present in the other transformants. The simplest explanation, then, for the lack of a phenotype in the 35S:ZF transformants is the low abundance of the mRNA of the expected size. This may have occurred because the truncated transcript is unstable, because the fusion transcript between the truncated CO gene and the CaMV polyadenylation sequence is unstable or because of the presence of a rearrangement in the T-DNA which affects the transcript size and its abundance.

The homeodomain transcription factor KNOTTED-1 from maize regulates vegetative development (Vollbrecht *et al.*, 1991). When the DNA binding domain and conserved proximal amino acids, the ELK-homeodomain (ELK-HD), was expressed under the control of the 35S promoter in transgenic tobacco no dominant negative phenotypes were observed (Meisel and Lam, 1996). One explanation was that the DNA binding domain, which was fused to the *uidA* reporter gene, was not properly folded, and therefore could not bind to DNA. Alternatively residues outside this domain may be required for the stability of the bound protein. Another explanation, then, for the lack of dominant negative phenotypes among transformants could be that even if proper translation of the modified co proteins occurs, the resulting proteins are unstable or not folded properly. When these experiments were performed, an antibody to the CO protein was not available to verify whether the stable mutant protein was present in the transformants.

Only a small number of transformants were obtained for the 35S:*co-3* and 35S:ZF constructs. Such a small number of transformants may have been biased against those with strong phenotypes so that only transformants in which the expression of the transgene was low or modified were recovered. Low expression of the transgene was observed for the 35S:ZF transformants. Expression of the modified genes under the control of an inducible system rather than a constitutive promoter, for example by fusing the modified co proteins

to the steroid-inducible rat glucocorticoid receptor would overcome this difficulty by permitting normal propagation of the transformants in the absence of induction. The uninduced transformants would also serve as a more stringent experimental control than untransformed wild type plants.

A total of 37 kanamycin resistant transformants were produced with 35S:*co-7*. Although only five of these were analyzed in detail in the T₃ generation in both LDs and SDs, many others were observed to flower earlier in the T₂ generation than wild type and the 35S:ZF and 35S:*co-3* transformants. The generation of a large number of transformants and the observation that many of them were slightly accelerated in flowering time compared to wild type suggested that the *co-7* protein under the control of the 35S promoter had some residual CO activity which was sufficient to promote flowering, although less efficiently than the wild type protein. The transformants containing this construct also weakly resembled the 35S:CO transformants (Igeño *et al.*, unpublished results) in the formation of a terminal pistil. Although the *co-7* allele is the second most severe in the allelic series and flowers later than the *co-8* mutant which is presumed to carry a null allele, it could still retain some function. However, this residual function of *co-7* has probably been dramatically enhanced by expressing the mutant protein from the strong 35S promoter. This would express the mutant protein at artificially high levels, in tissues and at developmental stages where it is not normally expressed.

Another possible explanation for the slight acceleration of flowering seen in the 35S:*co-7* transformants is that the CO protein may be negatively regulated by an interacting repressor protein, and that the overexpressed, defective *co-7* protein may titrate this repressor into a non-functional complex. This would remove the negative regulation from the wild type endogenous CO protein which may result in an enhancement of CO function and consequently earlier flowering plants.

Unlike the overexpression of *co-7*, overexpression of the *co-3* allele did not accelerate flowering. Therefore the overexpression of *co-3* cannot compensate for the loss of function of the putative DNA binding domain which is affected in *co-3*. However, overexpression of *co-3* did not cause a dramatic dominant negative phenotype either. The original model suggested that the *co-3* protein could not bind to DNA, but might interact with another factor, which, if *co-3* was present in excess, would be unavailable to interact with the wild type CO protein. Perhaps CO cannot interact with another factor if it is not bound to DNA, so that the wild type CO protein, present in the Landsberg *erecta* transformants, can promote flowering regardless of the presence of excess *co-3*.

Alternatively, if CO can only bind to DNA as a homo- or hetero dimer, this might explain why 35S:ZF transformants, if they contain a stable modified *co* protein, showed no mutant phenotype. A truncated CO protein lacking the conserved C-terminal domain may not be able to interact with its partner, and therefore cannot bind to DNA. If this is the case then the endogenous wild type CO protein will not be out-competed at the DNA binding sites.

Although none of the modified *co* constructs produced a dominant negative effect on flowering time, many long day-grown transformants had an increased number of petals (usually five or six instead of four) in their first or second flowers. The petal number phenotype was seen most frequently in the transformants carrying the 35S:ZF construct, although its low penetrance made it difficult to study. The phenotype is sometimes observed among wild type plants grown in long days, but was not observed in the Landsberg *erecta* controls grown alongside the transformants during these experiments. However, as the control population of Landsberg *erecta* plants was untransformed and had not been through selection in tissue culture the possibility of this phenotype being a tissue culture- or T-DNA- induced artefact cannot be ruled out.

As previously discussed in Chapter 2, an increase in petal and stamen number is a common phenotype in short day grown plants, and in long day grown *co* and other late flowering mutants. The incidence of increased petal number in these transformants and the increase in carpel number in the 35S:*CO* transformants (Igeño *et al.*, unpublished results) and some of the transformants carrying the *Brassica napus CO* orthologue (Chapter 6) suggests that *CO* may interact with regulators of meristem development such as *CLAVATA1*. Mutations in the *CLAVATA1* gene result in enlarged vegetative, inflorescence and floral meristems and an increase in floral organ number (Clark *et al.*, 1993). The *clavata1* mutants were originally identified as having an increase in carpel number (Koornneef *et al.* 1983) but subsequently were shown to have an increase in the number of all floral organ types and sometimes in the number of whorls. Weaker alleles tend to show an increase in carpel and stamen number and stronger mutant alleles show additional increases in petal and sepal number.

The analysis of dominant negative mutations of the MADS box protein AGAMOUS effectively demonstrates the systematic dissection of a protein into functional domains. Mizukami *et al.* (1996) constructed thirteen AGAMOUS derivatives based on the sequence homology to related genes which divides AGAMOUS into five distinct regions. They investigated the *in vivo* functions of each domain by analyzing the phenotypes of transgenic plants over-expressing these derivatives, and also used them *in vitro* to study DNA binding and dimerization. This strategy could be applied to CO, which can also be subdivided into a number of regions and recognizable structural domains: an N-terminal region, zinc finger 1, zinc finger 2, the putative transcriptional activator, the putative NLS, the highly conserved basic domain, the C-terminus and the less well conserved regions between some of these domains. The creation of a number of derivatives, where one or more of these regions is deleted might be effective in producing dominant negative phenotypes, but also

may help to analyze the function of individual subdomains. By overexpressing several different co derivatives at least some of them may be stably translated and folded properly.

The creation of a dominant negative mutation of the GATA-3 transcription factor in a transient expression system (Smith *et al.*, 1994) may also suggest one strategy for the design of modified CONSTANS proteins. N-terminal and C-terminal deletions of GATA-3, which left the zinc fingers intact could not suppress wild type GATA-3 function. The region between the two zinc fingers of GATA-3 is highly conserved to that of GATA-1 and contains a small basic motif. Mutagenesis of three basic residues (KRR) to AAA caused GATA-3 to display a dominant inhibition of transactivation of a reporter gene in the presence of wild type GATA-1, -2 or -3. The CO protein also contains a basic domain between the two zinc fingers, which is conserved between CO homologues and appears to be important for CO function as it is affected by a point mutation in the *co-2* allele. Site-directed mutations or small deletions within this domain may affect transactivation but leave DNA binding intact, leading to a dominant negative inhibition of wild type CO function.

Since the experiments described in this chapter were planned and performed the characterization of the CO protein has continued. In particular, the sequence similarities between CO and its homologues have allowed important regions of homology to be identified. The strategy used for creating dominant negative mutations could be improved in light of this information. A comprehensive series of deletions of specific domains from either end or from within the CO protein, site directed mutagenesis of particular residues and expression of isolated domains may be necessary in order to create dominant negative derivatives. These derivatives would also help to assess the function of the individual domains of CO, both *in vitro* and *in vivo*.

6 Investigation of the conserved function of a *Brassica napus* CO homologue

6.1 Introduction

Arabidopsis thaliana and *Brassica napus* (oilseed rape) are taxonomically closely related, both belonging to the Cruciferae, or Brassicaceae, family (Hedge, 1976). Unlike *Arabidopsis*, some members of the *Brassica* genus are of great agricultural importance, providing vegetables and spices for human consumption, animal fodder, and oils for culinary purposes and for use as industrial lubricants (Murphy, 1992). Genetic mapping studies in *Brassica* species are being carried out in order to map genes controlling agronomically important traits (Campos *et al.*, 1996; Lagercrantz *et al.*, 1996; Keith, 1995; Teutonico and Osborn, 1994). However, the complex, internally duplicated genomes of *Brassica* species makes the isolation of such genes difficult (Lagercrantz and Lydiate, 1995). On the other hand, *Arabidopsis*, with its small genome, high number of single-copy genes and low amount of repetitive DNA, is ideal for isolating genes using map-based cloning and chromosome walking. Genes isolated in this way from *Arabidopsis* could potentially be used to isolate homologues from their close relatives in the *Brassica* genus. Considerable variation in flowering time exists among members of the Brassicaceae and the manipulation of flowering time could have a significant impact on the cultivation of these crops. Currently, for example, *Brassica napus* is only grown in temperate latitudes because of the lack of very early flowering varieties, which would be required if it were to be cultivated in subtropical regions or the more northerly latitudes of Canada (Akbar, 1987; Thurling and Depittayanan, 1992; Murphy and Scarth, 1994).

The mapping of a number of QTLs (Quantitative Trait Loci) controlling flowering time in some *Brassica* species has identified genomic regions that contain homologues of the *CO* gene (Keith, 1995; Lagercrantz *et al.*, 1996). In *Brassica nigra*, *CO* homologues were located to two loci on linkage groups LG2 and LG8, close to or coincident with loci that influence flowering time (Lagercrantz *et al.*, 1996). The *CO* homologue on LG2 appeared to be coincident with a major flowering time QTL in *Brassica nigra*, controlling 53% of the total variation. In *Brassica napus*, *Arabidopsis CO* probes and other probes from the top arm of chromosome 5, detected a number of loci on several different linkage groups at which QTLs controlling flowering have been mapped (Keith, 1995; Osborn *et al.*, 1997). *CO* homologues might be expected at six loci in *Brassica napus* because it is an amphidiploid derived from the hybridization of *Brassica rapa* and *Brassica oleracea* and the genome of each of the progenitor species is believed to be triplicated (Lagercrantz and Lydiate, 1996). The *Brassica napus CO* homologues have been mapped to linkage groups N2 and N10, from the *Brassica rapa* component of the genome, and linkage groups N12 and N19 from the *Brassica oleracea* component (linkage group nomenclature according to Parkin *et al.*, 1995). Whether all of these *CO* genes have remained active and have a role in the promotion of flowering is addressed in this chapter.

The conservation of synteny (conservation of genetic linkage between the same genes in different species), collinearity (gene order) and gene sequence **between** *Arabidopsis thaliana* and members of the *Brassica* genus are being extensively investigated (Cavell *et al.*, 1998; Kowalski *et al.*, 1994; Lagercrantz *et al.*, 1996; Sadowski *et al.*, 1996; Scheffler *et al.*, 1997). This chapter is concerned with the conservation of the function of a *Brassica napus CO* homologue. The work described in the remainder of the introduction to this chapter was carried out by Dr. Laurian Robert, who provided a cloned *Brassica napus CO* gene to the author for complementation of the *co* mutant phenotype in *Arabidopsis thaliana*.

The *Brassica napus* *CO* homologues were isolated from genomic libraries constructed from the doubled-haploid lines N-o-1 and N-o-9 that show differences in their flowering time. N-o-1 is a facultative early flowering annual spring variety from Canada. N-o-9 is a late flowering biennial winter variety that has an absolute requirement for vernalization in order to flower. The *Arabidopsis* *CO* cDNA, amplified by PCR, was used as a probe against these libraries, and a number of genomic clones were isolated. From their restriction patterns, these genomic clones appeared to represent four loci, two from each early or late flowering line. The chosen representative clones were RFLP mapped and found to correspond to the *CO* locus on linkage group N10 from the *Brassica rapa* (A) genome and the homoeologous locus on linkage group N19 from the *Brassica oleracea* (C) genome. Both loci were isolated from the two lines screened. Since the libraries were made from doubled-haploid lines each locus is assumed to be homozygous for a particular allele. The *CO* loci on linkage groups N2 (A genome) and N12 (C genome) were not isolated from either of the two genomic libraries.

The four genes were sequenced and found to be very similar to each other, with the highest similarity seen between alleles from the same ancestral genome but derived from different lines. The representative allele chosen for complementation analysis in *Arabidopsis* was derived from the *Brassica napus* early flowering spring variety, N-o-1 on linkage group **N10** (*Brassica rapa* A genome), and is known as *Bn1CON10*.

The isolation and characterization of the four *Brassica napus* *CO* homologues by Dr. Laurian Robert, and the complementation of the *Arabidopsis* *co-2* mutation by the author has been accepted for publication (Robert *et al.*, 1998).

6.2 Sequence comparison of the *Brassica napus* CO homologue Bn1CON10, and *Arabidopsis thaliana* CO

The sequences of the *B. napus* CO homologue Bn1CON10 were compared to the *Arabidopsis* CO sequence. The DNA sequences of the ORFs were 83% identical but in the two highly conserved domains, discussed in Chapter 3, the identity was higher. The region encoding the zinc fingers was 89% identical and the region encoding the C-terminal basic domain was 91% identical. Upstream of the start of translation the homology was maintained at 66% for approximately 125bp, including a perfectly conserved sequence (TATTA) which may represent a TATA box. Further upstream the similarity decreased. Like the *Arabidopsis* CO gene, the Bn1CON10 allele, and the other alleles isolated from *Brassica napus*, have a single intron, and its position is conserved in all of the genes. The Bn1CON10 and CO predicted proteins were 74% identical but this increased to 91% over the zinc finger domain and 94% over the C-terminal basic domain. All the residues which, from the analysis in the introduction (Chapter 1) and Chapter 3 of this thesis, are thought to be important for CO function within these two domains are conserved in Bn1CON10. These include the putative zinc-binding cysteines, the residues which are affected in the seven alleles and the putative nuclear localization consensus. Interestingly, the sequence encoding the most highly negatively charged region of the candidate transcriptional activating sequence in *Arabidopsis* CO is deleted in the *B. napus* CO homologue. Figure 6.1 shows the *B. napus* and *Arabidopsis* CO amino acid sequences aligned.

6.3 Plasmid construction

Two plasmid constructs containing the Bn1CON10 allele were designed. The first contained the two ORFs separated by the intron, 2.5 kb of DNA upstream of the ATG and 2.2 kb after the TGA. 1.1 kb of DNA 5' to the start of translation was sufficient to drive expression of the *Arabidopsis* CO gene and therefore 2.5 kb was assumed to contain all the

cis elements necessary for regulating the expression of *Bn1CON10*. The second gene fusion contained the *Bn1CON10* ORFs driven by the *Arabidopsis CO* promoter, in case the *Brassica* promoter was not active in *Arabidopsis*.

6.3.1 The *Brassica napus CO* homologue expressed from its native promoter

This construct (called pBnCO) was made by Dr. Laurian Robert, who inserted the 6.1 kb *Eco* RI fragment from the genomic clone $\lambda Bn1CON10$ into the *Eco* RI site of the binary vector SLJ1711 (Jones *et al.*, 1992; Figure 6.2A and B). The fragment contains the entire *CO* coding region of allele *Bn1CON10* plus approximately 2.5 kb and 2.2 kb of 5' and 3' flanking region, respectively. The resulting plasmid, pBnCO, was introduced into *Agrobacterium tumefaciens* C58C1 pGV2260 by the author. The integrity of the construct was verified by restriction digestion and Southern analysis of plasmid DNA prepared from putative *Agrobacterium* transformants.

6.3.2 The *Arabidopsis CO* promoter with the *Brassica napus CO* coding sequence

To make this construct (called pApBnCO) a 2.3 kb *Hind* III fragment containing the *Bn1CON10* coding region plus approximately 70 bp and 0.9 kb of 5' and 3' flanking region, respectively, was subcloned from the genomic clone $\lambda Bn1CON10$ into pBluescript (Figure 6.2A and C). The *Hind* III site just upstream of the start of translation is present also in the *Arabidopsis CO* gene, and has been used in the construction of a number of transcriptional fusions between *CO* and foreign promoters (Simon *et al.*, 1996) The 70 bp of 5' sequence from the ATG up to this *Hind* III site, plus an additional 60 bp is highly conserved between the *Arabidopsis CO* gene and the four *Brassica napus CO* homologues sequenced so far (Robert *et al.*, 1998). The *Arabidopsis CO* promoter was isolated as a 2.1 kb *Hind* III fragment from an *Arabidopsis CO* genomic clone, p486, kindly provided by

Bn101 Co	M F K Q E S N N I G M L K O E S N D I G	S E E N N T G P R A S G E N N R A . R P	C D T C G S T I C T C D T C R S N A C T	V Y C H A D S A Y L V Y C H A D S A Y L	C N S C D A Q V H S C M S C D A Q V H S	A N R V A S R H K R A N R V A S R H K R	60 59
Bn101 Co	V R V C E S C E R A V R V C E S C E R A	P A A F M C E A D D P A A F L C E A D D	V S L C T A C D L E A S L C T A C D S E	V H S A N P L A R R V H S A N P L A R R	H Q R V P V V P I T H Q R V P I L P I S	G N S C S S L A T A G N S F S S M T T T	120 119
Bn101 Co	N F . . . T T V T E H E Q S E K T M T D	P E K R V V L V Q E P E K R L V V D Q E D A K E G E E G D K A K	E T A S W L F F P K N E V A S W L F P . N	S D N H N N N N Q N S D K . N N N N Q N	N E L L F S D D Y L N G L L F S D E Y L	170 177
Bn101 Co	D L A D Y N S S M D N L V D Y N S S M D	Y K F T G Q Y N Q P Y K F T G E Y . . .	T Q H K Q D C T V P S Q H Q Q N C S V P	E K N Y G G D R V V Q T S Y G G D R V V	P L Q L E E T R G N P L K L E E S R G H	L H H K Q H . . . Q C H N Q Q N F Q F	225 234
Bn101 Co	N I T Y G S S G S H N I K Y G S S G T H	Y N N N G S I N H N Y N D N G S I N H N	A Y N P S M E T D F A Y I S S M E T G V	V P E Q T A P D K T V P E S T A C V T T	V S H P K T H K G K A S H P R T P K G T	I E K L P E P L I Q V E Q Q P D P A S Q	286 294
Bn101 Co	I L S P M M I T V T Q L S P M	D R E A R V L R Y R D R E A R V L R Y R	E K K K R R K F E K E K R K T R K F E K	T I R Y A S R K A Y T I R Y A S R K A Y	A E R R P R I N G R A E I R P R V N G R	F A K I S E T E V E F A K . R E I E A E	341 353
Bn101 Co	D Q E Y N T M L M Y E Q G F N T M L M .	Y D T G Y G I V P S Y N T G Y G I V P S	F Y G Q K 365 F 373				

Figure 6.1 Prettybox comparison of the predicted protein sequence of the *Brassica napus* Bn1CON10 allele (Bn101) with *Arabidopsis* CO (Co).

Kate Dixon (Figure 6.2D). The promoter was cloned into the conserved *Hind* III site upstream of the *Bn1CON10* ORF and orientated by restriction mapping (Figure 6.2C). This was accomplished using a step-wise approach, because of the presence of a *Hind* III site at the 3' end of the *Bn1CON10* fragment. The *Hind* III site at the 3' end was temporarily removed within an *Xba* I fragment while the promoter was added to the coding sequence, and replaced in the correct orientation after all *Hind* III cloning steps were completed. The *Arabidopsis* promoter:*Brassica napus* ORF fusion was moved as a 4.5 kb *Cla* I/*Sac* I fragment into the binary vector SLJ1711, and introduced into *Agrobacterium tumefaciens* as previously described.

6.4 Introduction of the constructs into the *co-2* mutant

The two constructs were introduced into the *Arabidopsis thaliana co-2* mutant by a root transformation method adapted from Valvekens *et al.* (1988). The *co-2* mutant used was marked with the *tt4* (*transparent testa*) mutation which maps approximately 3.4 cM from *co* and causes a yellow seed phenotype (Koornneef *et al.*, 1990) so that any early flowering transformants caused by complementation of *co-2* would be distinguishable from wild type brown seeded plants.

The BnCO construct and the ApBnCO construct yielded seven and eight independent kanamycin-resistant transformants, respectively. These T₁ plants were allowed to self-fertilize, their T₂ progeny were plated onto GM with kanamycin at 50mg/L and the number of kanamycin resistant:sensitive seedlings recorded to determine the number of loci at which the T-DNA had inserted (Table 6.1). The seeds of the transformants were yellow, verifying the presence of the *tt4* marker and excluding the possibility that they were derived from contaminating wild type plants. The kanamycin resistant T₂ seedlings were transferred to the greenhouse and self-fertilized. The resulting T₃ seeds were harvested and sown on

kanamycin-containing medium in order to identify individuals homozygous for the T-DNA insertion for each independent transformant.

6.5 Identification of transgenic plants homozygous for the T-DNAs

Approximately 100-150 T₃ seeds from each of five T₂ families for each transformant were sown on GM/Km and the ratio of resistant:sensitive seedlings recorded (Table 6.1). T₃ families that were all resistant to kanamycin and therefore derived from T₂ plants homozygous for the T-DNA were identified for all seven BnCO transformants, and for five of the ApBnCO transformants, and analyzed further in controlled flowering time experiments (Table 6.2)

6.6 Phenotypic analysis of the transformants under long and short days

For each transformant the flowering times of 20 plants homozygous for the T-DNA were measured under LDs and SDs and compared with the flowering times of Landsberg *erecta* and the *co-2 tt4* mutant. As described previously flowering time was scored as the number of days from sowing to the first appearance of buds in the centre of the rosette, and the total number of leaves produced on the primary shoot of the plant.

6.6.1 Flowering time of transformants carrying BnCO

Under LDs, the BnCO construct not only complemented the *co-2* mutant phenotype and restored wild type flowering time, but all transformants flowered earlier than the wild type control (Figure 6.3A; Table 6.2). This was most apparent in lines which were predicted to contain the T-DNA at multiple loci based on the segregation of kanamycin resistance. Under SDs the acceleration of flowering of the transformants was more evident as both wild type Landsberg *erecta* and the *co-2* mutant flower very late under these conditions. All transformants, including those predicted to contain the T-DNA at one locus, flowered

Construct	Transformant	K ^{mr} :K ^{ms} ratio in T ₂ generation	T3 family	K ^{mr} :K ^{ms} ratio in T ₃ generation	Comments
BnCO	Tn2a	3.75:1	Tn2a-3	2.9:1	
			Tn2a-5	1:0	Hom
	Tn3a	1.5:1	Tn3a-3	1:0	Hom? All T ₃ families 100% K ^{mr}
	Tn17	3.4:1	Tn17-2	2.7:1	
			Tn17-5	1:0	Hom
	Tn18a	6.0:1	Tn18a-3	5.6:1	
			Tn18a-4	1:0	Hom
	Tn34	2.0:1	Tn34-4	1:0	Hom? All T ₃ families 100% K ^{mr}
	Tn36	2.46:1	Tn36-2	2.4:1	
			Tn36-4	1:0	Hom
	Tn91	12.5:1	Tn91-6	7.0:1	
			Tn91-7	1:0	Hom
ApBnCO	Tn406	nt	Tn406-3	2.6:1	
			Tn406-4	1:0	Hom
	Tn410	3.7:1	Tn410-2	3.1:1	
			Tn410-1	1:0	Hom
	Tn412	nt	Tn412-3	3.2:1	
			Tn412-2	1:0	Hom
	Tn413	nt	Tn413-1	3.2:1	
			Tn413-2	1:0	Hom
	Tn418	nt	Tn418-3	4.0:1	
			Tn418-5	1:0	Hom

Table 6.1 The Kanamycin segregation analysis for the transformants in the T₂ and T₃ generations and the identification of families homozygous for the T-DNA/s to be used in further analysis of flowering time.

Transformant	Long Days		Short Days	
	Days to induction of flowering	Total leaf number	Days to induction of flowering	Total leaf number
BnCO in <i>co-2</i>				
Tn2a-5	18.7±0.2	6.9±0.1	35.6±0.5	23.7±0.4
Tn3a-3	16.8±0.5	5.4±0.1	29.9±1.1	11.1±0.6
Tn17-5	15.9±0.7	5.9±0.1	37.7±1.3	17.0±2.0
Tn18a-4	17.8±0.2	6.4±0.1	29.3±0.7	13.4±0.8
Tn34-4	18.5±0.4	6.8±0.2	34.0±1.0	13.2±0.8
Tn36--4	18.3±0.3	6.8±0.1	34.8±1.0	12.9±0.8
Tn91-7	17.4±0.6	5.2±0.2	23.4±0.8	7.7±0.5
<i>co-2 tt4</i>	33.1±0.9	19.1±0.6	48.8±0.8	36.7±1.1
Wild type <i>Ler</i>	19.8±0.9	8.2±0.2	60.3±1.2	48.9±0.8
ApBnCO in <i>co-2</i>				
Tn406-4	30.2±1.2	15.9±0.6	NT	
Tn410-1	26.4±0.4	12.8±0.3		
Tn412-2	27.6±0.8	13.2±0.5		
Tn413-2	27.7±0.8	13.0±0.6		
Tn418-5	25.9±0.5	13.0±0.5		
<i>co-2 tt4</i>	30.9±2.0	14.4±0.9		
Wild type <i>Ler</i>	20.1±0.5	7.2±0.3		

Table 6.2 Flowering time and leaf number analysis for the transformants in the *co-2 tt4* mutant background, in LDs and where applicable, SDs, compared to untransformed *co-2 tt4* and wild type *Landsberg erecta*

A



B

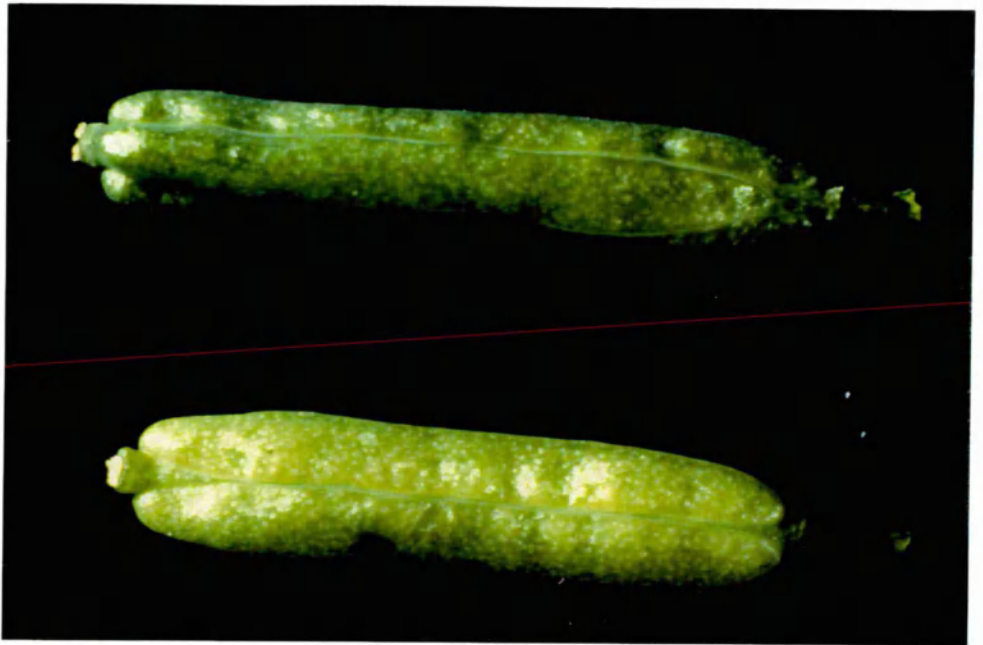


Figure 6.3 A) From the left, a wild type Landsberg *erecta* plant, a *co-2 tt4* mutant and a *co-2 tt4* mutant carrying the *Brassica napus CO* homologue, *Bn1CON10* (the BnCO construct, transformant Tn2a). All plants were grown in LDs and are 28 days old. B) Plants as in A, but grown in SDs. All plants are 42 days old.

earlier and with fewer leaves than both control populations under SDs (Figure 6.3B; Table 6.2). However, the early flowering phenotypes of lines containing T-DNAs at multiple loci were particularly dramatic. For example, transformant 91 (Tn 91), flowered in SDs with an average of 7.8 leaves after 23.4 days while the untransformed *co-2 tt4* control flowered with 36.7 leaves after 48.8 days. The Tn 91 family chosen for flowering time analysis appeared to be still segregating for one T-DNA which caused very early flowering, because some individuals in this SD grown population flowered with a total of 3 or 4 leaves.

6.6.1.1 Phenotypes other than early flowering conferred by pBnCO

Although all the BnCO transformants flowered earlier than wild type in both LDs and SDs there was a gradation of flowering time phenotypes, and the most extremely early individuals displayed other phenotypes which are characteristic of plants in which *CO* is expressed from the CaMV 35S promoter. Transformants which carry 35S:*CO*, as well as flowering early in both LDs and SDs form fewer flowers than wild type, terminate in the formation of a pistil, and show alterations in flower morphology, particularly to the gynoecium which can be comprised of three or four carpels rather than the usual two (Igeño *et al.*, unpublished results). These phenotypes are also found in the earliest flowering BnCO transformants (particularly individuals derived from Tn 91 and Tn 3a) in both LDs and SDs. Figure 6.4A shows an altered BnCO gynoecium containing three fused carpels, compared to a wild type gynoecium. Several early flowering individuals from a number of BnCO transformed lines also showed an alteration in phyllotaxy on the flowering stem. Flowers were no longer produced in the spiral pattern characteristic of *Arabidopsis*, but were randomly distributed along one side of the stem (Figure 6.4B and C). This phenotype was not observed in the 35S:*CO* transformants.

A

1 mm

B

1 cm

C

Figure 6.4 Examples of phenotypes observed in some BnCO transformants.

A) At the top, an altered gynoecium from a Tn 91 individual, comprised of three fused carpels compared to a wild type gynoecium below, comprised of two.

B) A flowering stem from a Tn 91 individual showing an altered phyllotaxy, compared to a wild type stem (**C**) with normal spiral phyllotaxy.

6.6.2 Flowering time of transformants carrying ApBnCO

All transformants carrying the *Brassica napus* CO ORF driven by the *Arabidopsis* CO promoter flowered as late as untransformed *co-2 tt4* under LDs, so this construct was not able to complement the *co* mutant phenotype (Figure 6.5; Table 6.2). The integrity of the construct in each of the five transformants analysed was checked by restriction digestion of DNA extracted from the transformed plants, Southern blotting and hybridization to CO and T-DNA probes (data not shown). No rearrangements which might compromise CO expression were detected at this level of analysis.

6.7 Analysis of the expression of the transgenes by Northern hybridization and RT-PCR

To confirm that the ApBnCO transgene was expressed and to compare the abundance of the CO transcript in the ApBnCO and BnCO transformants, the expression of the transgenes in each transformant was analyzed. As the BnCO construct complemented the *co-2* mutation and caused phenotypes associated with over-expression of CO, Northern analysis was first attempted, in case the BnCO transcript in the transformants was present at levels detectable by this method. In wild type plants, the CO transcript was too rare to be detected by Northern analysis (Putterill *et al.*, 1995). RNA was extracted from triplicate batches of 10 day old LD grown seedlings of the following genotypes: each BnCO transformant, wild type Landsberg *erecta*, *co-2* mutant and, as a positive control, the 35S:CO transformant Tn5, provided by Maribel Igeño. The resulting Northern blot was probed with a mixed DNA probe containing the 6.1 kb *Brassica napus* BnCO fragment and the *Arabidopsis* CO cDNA (Figure 6.6). No transcript specific to the BnCO transformants was detected but a very abundant transcript was present in the 35S:CO samples. A faint transcript of the same size as the 35S driven CO transcript in the positive control plants was present in all samples including the untransformed controls but this was probably a



Figure 6.5 From the left, a wild type Landsberg *erecta* plant, a *co-2 tt4* mutant and two *co-2 tt4* mutants, each carrying the *Arabidopsis* promoter:*Brassica napus* ORF construct, ApBnCO, which did not complement the *co* mutant phenotype. All plants were grown in LDs and are 28 days old.

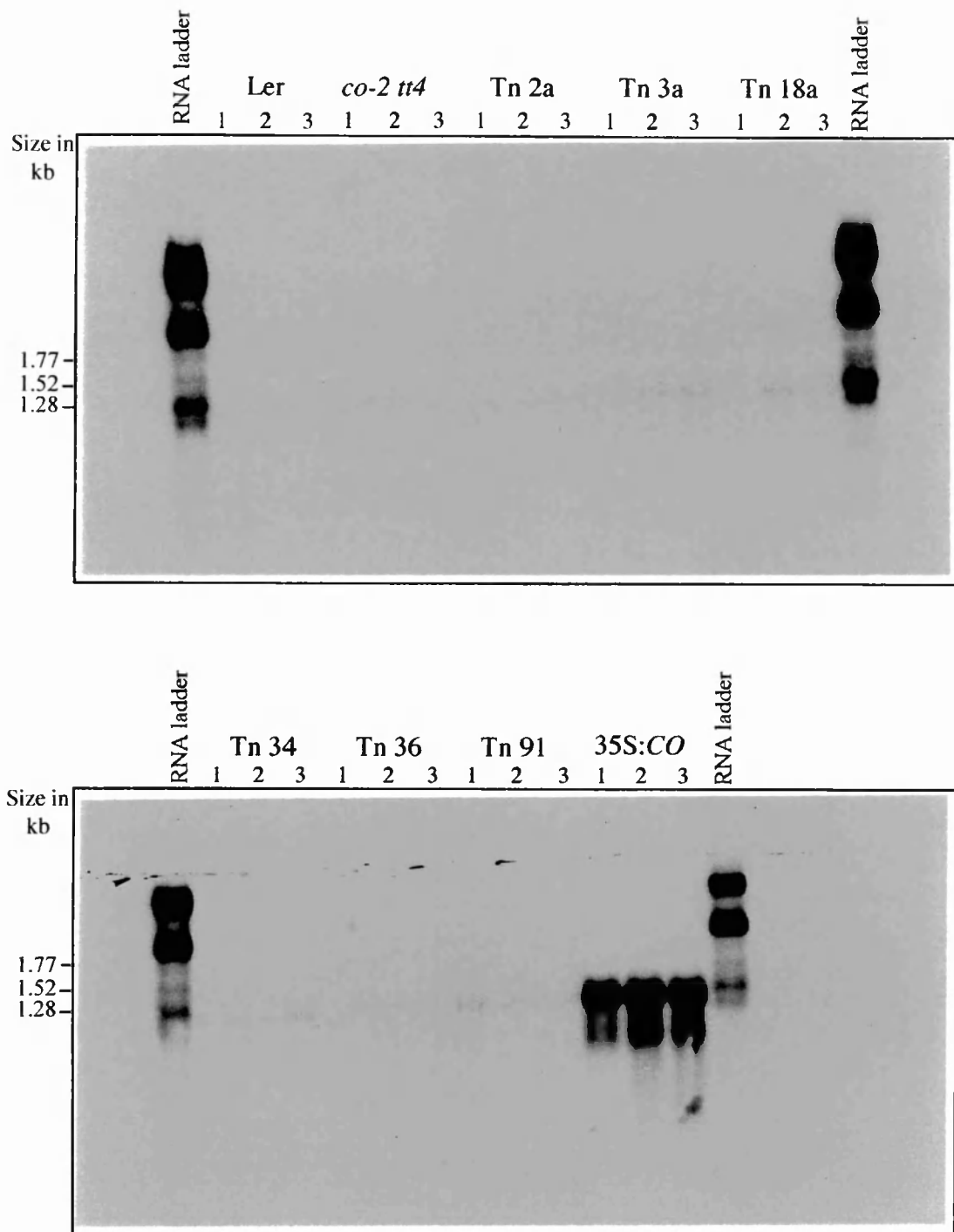


Figure 6.6 Northern blots of RNA extracted from the BnCO transformants. RNA from untransformed wild type Landsberg *erecta* and *co-2* mutants, and from plants carrying the CaMV 35S:CO construct were included as controls. Three independent RNA samples were loaded for each transformant or control as the ethidium bromide-stained gel also served to check the concentration of RNA for duplicate RT-PCR reactions (Figure 6.7). The blot was probed with the CO cDNA and the 6.1 kb *EcoR* I fragment containing the whole *BnICON10* gene.

highly expressed *CO* homologue like *COL1*, which cross hybridizes with *CO* in both Northern and Southern analysis unless a *CO*-specific probe is used (Putterill *et al.*, 1997). The ApBnCO transformants were not analyzed by Northern hybridization because the BnCO ORF in this construct is expressed from the *Arabidopsis* promoter and consequently the *CO* mRNA in these lines was assumed to be undetectable by this method.

To analyze the wild type *CO* transcript an RT-PCR approach was taken (Putterill *et al.*, 1995). This technique was therefore also used to look at expression of the transgenes in BnCO and ApBnCO transformants.

To analyze the expression in *Brassica napus* of four *CO* homologues, Laurian Robert designed PCR primers which were allele-specific. A pair of primers designed to only amplify the *Bn1CON10* allele were kindly provided to the author. One of these primers, bnco9, annealed to exon sequences on either side of the single intron present in the four *Brassica napus* *CO* homologues, to prevent amplification of contaminating DNA and the second primer, bnco7, annealed at the 5' end of the first exon and was specific for the *Bn1CON10* allele (Figure 6.7b; Chapter seven, Materials and Methods). These specific primers were able to amplify a fragment of the expected size (580 bp) from cDNA prepared from the RNA of both the BnCO and ApBnCO transformants. This transcript hybridized to a *CO* probe (Figure 6.7a). No major differences in transcript level were detected between transformants carrying the same construct, even for lines thought to contain multiple copies of the complementing BnCO T-DNA. However, the transcript was reproducibly more abundant in the BnCO transformants than in the ApBnCO transformants.

An additional fragment of approximately 750 bp was also amplified from all transformants, but was more abundant in the ApBnCO transformants than in the BnCO transformants, and in the ApBnCO transformants was more abundant than the correct sized fragment. The nature of this extra fragment is as yet unknown, however it was able to hybridize to a *CO*

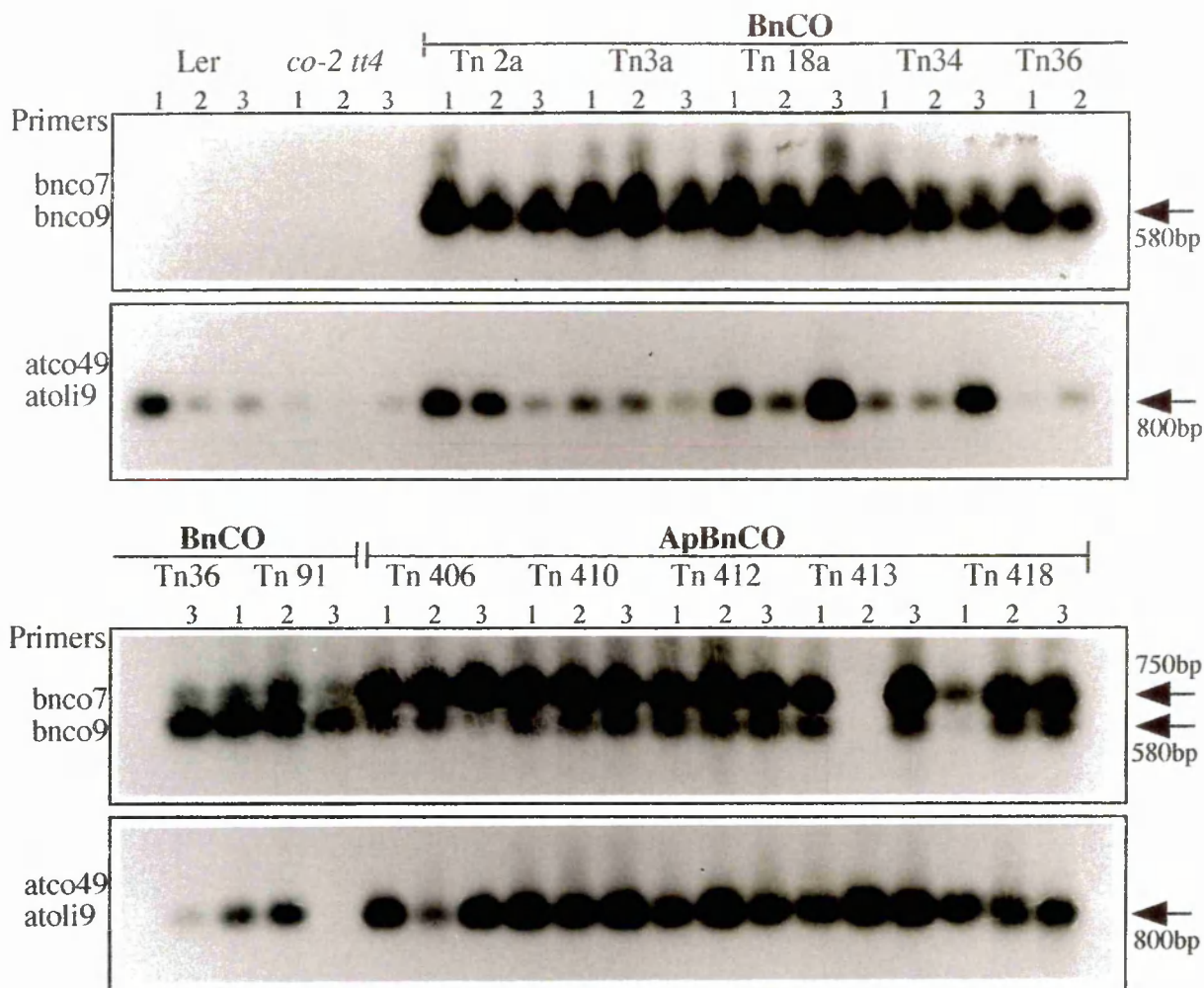


Figure 6.7a Southern blots of RT-PCR products amplified from RNA extracted from the BnCO and ApBnCO transformants and untransformed control plants. RNA was extracted from three batches of plant tissue for each transformant or control and used in independent reverse transcription and PCR reactions so that an occasional negative result, probably due to poor RNA quality, could be discounted. The blot was probed with a mixed DNA probe containing the *Arabidopsis CO* cDNA and the 6.1 Kb *Eco* RI *Brassica Bn1CON10* fragment.

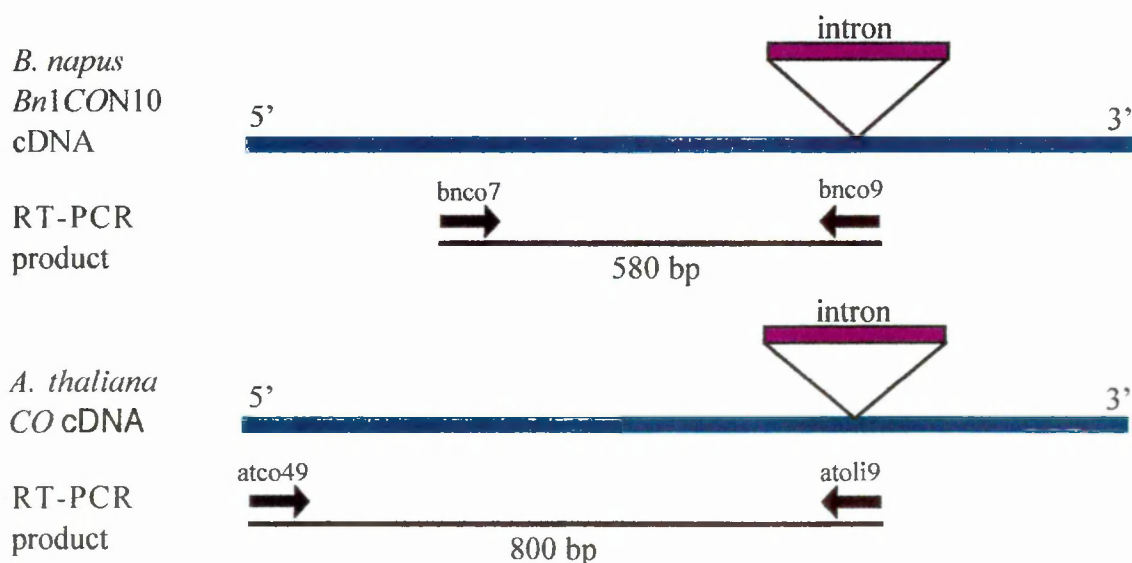


Figure 6.7b The RT-PCR strategy used to monitor the expression of the transgenes in the transformants carrying the *Bn1CON10* allele.

probe. No transcript was amplified using the *Bn1CON10* primers from the untransformed controls, confirming that the primers were specific to the transcript of the *B. napus* gene, and RT-PCR reactions using *Arabidopsis CO*-specific primers amplified a transcript from both transformed and untransformed plants as expected.

6.8 Discussion

A *CO* homologue from the *Brassica rapa* genome in a Canadian early flowering spring variety of *Brassica napus* was introduced into the *Arabidopsis co-2* mutant and demonstrated to complement the *co-2* mutation. This indicates that the necessary *cis* regulatory elements required for *CO* expression are contained within the 6.1 kb fragment used for complementation. The *B. napus* promoter must be recognized by the *Arabidopsis* transcriptional machinery and the protein encoded by the gene must be able to interact correctly with the factors necessary for CO function.

The *Bn1CON10* allele not only corrected the late flowering *co-2* mutation, but all seven independent BnCO transformants analyzed flowered earlier than wild type under both LDs and SDs. The segregation of kanamycin resistance in the T₂ families derived from some of the transformants suggested that they contain the transgene at multiple loci. The earliest flowering line, Tn 91, for example, had a kanamycin resistant:sensitive ratio of 12.5:1 in the T₂ generation and 7.0:1 in a T₃ family which would be consistent with the segregation of two loosely linked T-DNA copies. One of these copies appeared to be segregating in the T₃ family chosen for flowering time analysis because some extremely early flowering individuals with a total leaf number of three were observed in the SD-grown population (wild type, 48.9 leaves). Two other lines, Tn 3a and Tn 34 also had an unusual kanamycin segregation ratio, where all T₃ families tested appeared to be 100% kanamycin resistant. This could be coincidental, or could also suggest that these two lines contain many copies of the T-DNA. To estimate the number of T-DNAs, and identify homozygous families, the

segregation of kanamycin resistance would have to be followed in a larger population of seedlings or kanamycin resistant plants could be backcrossed to wild type to reduce the copy number. However, at least three of the transformed lines, Tn 2a, Tn 17 and Tn 36, were predicted to contain the T-DNA at only one locus and these lines also flowered earlier than wild type in both LDs and SDs. It is possible that these lines also contained multiple copies of the T-DNA tandemly arranged at one locus. Alternatively the position of integration of the T-DNA into the genome may be significant because the *CO* transcript is present at an extremely low level in wild type plants and consequently the presence of other regulatory elements close to the site of integration of the T-DNA may affect the expression of the transgene and therefore cause earlier flowering.

Introduction of the *Arabidopsis CO* gene into *co-2* mutants produced similar results to those obtained with the *Brassica CO* homologue (Putterill *et al.*, 1995). Several *co-2* transformants carrying a cosmid that contained the *Arabidopsis CO* gene flowered earlier than wild type in both LDs and SDs. Some of these were predicted to contain the T-DNA at several loci based on the kanamycin segregation ratios and on Southern blot analysis. One transformed line, when analyzed in SDs, contained individuals which flowered with five leaves (wild type flowered with 18.9) indicating that at least one T-DNA in this line was causing very early flowering. The results from the complementation with the *Arabidopsis CO*-containing cosmids were the first indication that *CO* could promote earlier flowering than wild type in a dosage-dependent manner, and that in SDs the amount of *CO* is limiting. Subsequently, expressing *CO* under the control of the strong CaMV 35S promoter (Igeño *et al.*, unpublished results) confirmed this observation. The transformants carrying this fusion are practically day neutral, flowering earlier than wild type after producing a total of approximately five leaves in both LDs and SDs.

In wild type plants the *Arabidopsis CO* transcript is present at very low levels (Putterill *et al.*, 1995) and *co* heterozygotes flower later than wild type (Chapter 2) suggesting that a

two-fold reduction in the amount of *CO* mRNA delays flowering. The transcript from the *Brassica CO* homologue in the *Arabidopsis* transformants could be detected by RT-PCR but not by Northern analysis, even for lines containing multiple T-DNAs which caused some individuals to flower as early as plants carrying the 35S:*CO* construct. This observation is in agreement with the idea that the level of expression of *CO* is critical in determining flowering time and that above a certain threshold of *CO* transcript abundance the response to *CO* is saturated (Putterill *et al.*, 1995).

The observation that transformants carrying *Bn1CON10* flower earlier than wild type may also be explained by differences in the *Brassica CO* protein, which is 74% identical at the amino acid level to the *Arabidopsis CO* protein, or to differences in the controlling regions of the gene. Upstream of the start of translation the two genes are 67% identical for approximately 130 bp, but after this the similarity decreases. Although in *Arabidopsis* the transcript of the *Brassica CO* homologue is very rare, it is possible that differences in the promoter might affect transcript abundance, location or developmental stage of expression, which might in turn affect flowering time.

Transformants carrying the *Arabidopsis CO* gene under the control of the CaMV 35S promoter (35S:*CO*) showed a number of modifications in flower and shoot morphology (Igeño *et al.*, unpublished results). These characteristic modifications, caused by increased and/or ectopic expression of *CO*, were also observed among transformants carrying multiple copies of the *Brassica CO* homologue. The transformants formed fewer flowers than wild type and these flowers were sometimes altered, particularly the gynoecium which was often composed of three or more fused carpels instead of the usual two. Another phenotype seen particularly in the earliest flowering *BnCO* transformants but not in the 35S:*CO* transformants was a disruption of the spiral phyllotaxy of the shoot, so that flowers were distributed along one side of the stem.

The *Bn1CON10* gene expressed from its own promoter was able to complement the *co-2* mutation in *Arabidopsis*. However, when the open reading frame of this gene was expressed from the *Arabidopsis* promoter (the ApBnCO construct) the transformants flowered as late as the untransformed *co-2* mutant controls. This could not be explained by large-scale rearrangements of the T-DNA because restriction digested DNA extracted from the transformants and probed with *CO* and T-DNA probes detected no rearrangements in the transgene. The transformants were then analyzed by RT-PCR to establish whether a *Bn1CON10* transcript was present. Transformants carrying either the BnCO construct or the ApBnCO construct were examined. Although the RNA concentrations were carefully measured, this RT-PCR experiment was not quantitative. However in duplicate experiments the *Bn1CON10* cDNA was amplified from RNA extracted from BnCO transformants and the abundance of the product suggested that the *Bn1CON10* RNA was slightly more abundant than in ApBnCO transformants. To verify this difference in transcript abundance, and to test whether it might account for the difference in ability to complement the *co-2* mutation the *Bn1CON10* transcript present in both sets of transformants should be analyzed more thoroughly by quantitative RT-PCR. If the transcript is more abundant when the *Bn1CON10* allele is under the control of the *Brassica*, rather than the *Arabidopsis*, *CO* promoter perhaps this increase in abundance is able to compensate for differences in the *Bn1CON10* protein, which might be less efficient at promoting flowering than the native *Arabidopsis* *CO* protein.

A transcript of the expected size was present in all transformants, however an extra *CO*-hybridizing fragment was also amplified. In subsequent amplifications in which the PCR conditions were altered, this fragment disappeared from the BnCO samples, but was still present in the ApBnCO samples. The nature of this extra fragment needs to be investigated further by DNA sequencing. It is not clear whether the *Bn1CON10*-specific primers amplified a fragment from contaminating DNA, whether one or both primers can anneal

elsewhere on the *Bn1CON10* cDNA, or whether this larger fragment is an incorrectly spliced *Bn1CON10* transcript, and whether this might also account for the lack of complementation conferred by this construct.

In addition to the five ApBnCO transformants used during this analysis several other transformants were generated which were not analyzed. These could be screened to identify lines which have a higher *Bn1CON10* mRNA abundance either due to the presence of multiple copies of the T-DNA or to position effects, and the effect of this increased abundance on the flowering time of the *co-2* mutant studied.

The *Brassica* protein was 74% identical to the *Arabidopsis* CO protein, but as expected the highest homology was found in the two domains which are conserved in a number of other CO homologues from *Arabidopsis* and other species (Chapter three). The zinc fingers were 91% identical and the C-terminal basic domain showed even higher homology with 94% identity. If CO is involved in the regulation of transcription it might be expected to contain a transcriptional activation domain. One candidate for this is an acidic region from residues 124 to 150, which contains a highly negatively charged cluster from 138 to 147. This cluster is absent in the *Brassica* CO homologue. It is not known whether this acidic domain in the *Arabidopsis* CO protein is involved in transcriptional activation of genes *in vivo*, but the absence of this domain from the *Brassica* CO protein might suggest a number of possibilities. The first is that it is not important for CO function. Another possibility is that it acts as a transcriptional activator in the *Arabidopsis* CO protein, but in the *Brassica* protein another acidic domain is used to activate transcription. Alternatively the acidic residues which are still present in this region may be able to function without the highly acidic subdomain. It would be interesting to investigate whether the *Brassica* CO protein, like the *Arabidopsis* CO, can activate the transcription of reporter genes in yeast, and whether this or other domains are responsible.

The *Brassica napus* and *Arabidopsis* CO genes show a high degree of sequence conservation, consistent with their evolutionary relatedness (Cavell *et al.*, 1998). Other homologous gene sequences analyzed so far show an average of 87% identity at the DNA level, and 86% at the amino acid level (Lydiat *et al.*, 1993; Cavell *et al.*, 1998). The amino acid sequence similarity between the *Arabidopsis* and *Brassica napus* CO homologues is lower, at 74%. This lower level of similarity possibly reflects the nature of the genes compared so far between the two species, many of which are essential genes involved in metabolic processes rather than transcription factors like CO which are involved in developmental regulation.

The first published example of complementation of an *Arabidopsis* mutation with a *Brassica napus* gene has already been mentioned in the general introduction (Chapter 1; Arondel *et al.*, 1992). It was one of the first examples of chromosome walking in *Arabidopsis*, in which the genomic clones covering the region of the *Arabidopsis* genome thought to contain an Omega-3 fatty acid desaturase gene were used to isolate cDNAs from a *Brassica napus* library made from RNA extracted from seeds. An equivalent *Arabidopsis* library made from developing seeds was not available. The *Brassica napus* cDNA was introduced into the *Arabidopsis fad3* (fatty acid desaturase 3) mutant and successfully complemented the mutant phenotype.

Another example of functional complementation of an *Arabidopsis* mutation with a *Brassica napus* homologue involves another flowering time gene, *FCA*. In *Arabidopsis* this gene is involved in promoting flowering through the autonomous flowering pathway, and mutations at this locus, which cause late flowering in both LDs and SDs, can be corrected by vernalization (Macknight *et al.*, 1997; Chapter 1). The *Brassica napus* *FCA* allele from the *Brassica oleracea* (C) genome, isolated from the late flowering line N-o-9, was

introduced into the *Arabidopsis fca* mutant, and wild type flowering time was restored to the transformants (R. Macknight, unpublished results).

So far only the *Bn1CON10* allele from the *B. rapa* (A) genome has been used to complement the *Arabidopsis co-2* mutation. Three other *CO* alleles have been cloned from *Brassica napus*: the homoeologous allele from the *B. oleracea* (C) genome in the same early flowering line, and the equivalent A and C genome alleles on the same linkage groups in the late flowering line (Robert *et al.*, 1998). In addition to the *CO* alleles on linkage groups N10 and N19, there are also alleles on linkage groups N2 and N12. These have not yet been cloned as they appear to be less well conserved between *Arabidopsis* and *Brassica* and are only detected at very low stringency in Southern analysis using the *Arabidopsis CO* gene as a probe (Robert *et al.*, 1998). The four *CO* homologues cloned so far appear to be expressed in the *Brassica napus* lines from which they were derived, and it would be interesting to test whether the other three cloned alleles, and the less well conserved alleles on linkage groups N2 and N12, from both early and late flowering lines, can complement the *Arabidopsis co* mutation. This would help to determine whether all alleles encode a functional CO protein.

7 Materials and Methods

7.1 Materials

7.1.1 Plant Materials

Arabidopsis thaliana var. Landsberg *erecta* obtained from Professor Maarten Koornneef (University of Wageningen).

Arabidopsis thaliana var. Columbia obtained from Professor Elliot Meyerowitz (Caltech) via Dr. Clare Lister.

co-1 allele obtained from Professor George Rédei (University of Missouri) via Professor Maarten Koornneef.

co-2 and *co-3* alleles obtained from Professor Maarten Koornneef via Dr. Caroline Dean.

co-4, *co-5*, *co-6* and *co-7* alleles obtained from Professor Maarten Koornneef. All *co* alleles are in a Landsberg *erecta* background except *co-1* which is in Landsberg.

tt4 mutant (Landsberg *erecta* background) obtained from Professor Maarten Koornneef.

co-Dfx41-1, *Dfx-1*, -2, -4, -5, -6, -7, -15, -16 and *Dfa-2* to 15 obtained from Dr. Igor Vizir (University of Nottingham).

7.1.2 Bacterial and Yeast Strains

Escherichia coli

DH5 α *supE44*, F⁻, *endA1*, *hsdR17*, *thi-1*, Δ *lacU169*[ϕ 80 *lacZ* Δ M15] *recA1*,
gyrA96, *relA1* (Hanahan, 1983)

HB101 F⁻, *hsd S20* (*r_B*⁻, *m_B*⁻), *supE44*, *ara14*, λ ⁻, *galK2*, *lacY1*, *proA2*, *rpsL20*, *xyl-5*, *mtl-1*, *recA13* (Raleigh and Wilson, 1986)

JM101 *supE*, *thi-1*, $\Delta(lac-proAB)$, F' [*traD36*, *proAB*, *lacI*^qZ Δ M15] (Messing, 1979)

XL1-BlueMR $\Delta(mcrA)$ 182, $\Delta(mcrCB-hsdSMR-mrr)$ 172, *endA1*, *supE44*, *thi-1*, *recA*, *gyrA96*, *relA1*, *lac*, λ^- (Bullock *et al.*, 1987)

Agrobacterium tumefaciens

C58C1 A Rifampicin resistant derivative of C58 cured for pTiC58 (Zambryski *et al.*, 1983)

Yeast

AB1380 *Saccharomyces cerevisiae* strain AB1380 (*mat-a*, *ade2-1*, *can1-100*, *lys2-1*, *trp1*, *ura3*, *his5* [psi+]) (Burke *et al.*, 1987)

7.1.3 Plasmids

pBluescript II

SK (+/-) a plasmid derived from pUC19 (Stratagene USA)

pSLJ1711 A binary vector based on the tetracycline resistant octopine Ti plasmid pRK290. The T-DNA carries a neomycin phosphotransferase gene (*NPTII*) under the control of the CaMV 35S promoter and a polylinker derived from pBluescript (Jones *et al.*, 1992). Obtained from Dr. Jonathan Jones.

pRK2013 Helper plasmid which carries the *incP* conjugation system to allow transfer of binary vectors to *Agrobacterium* (Ditta *et al.*, 1980).

pGV2260 A disarmed Ti plasmid in which the entire T-DNA region is substituted by pBR322. The *vir* region is intact and will act *in trans* to transfer the T-DNA from the binary vector to the plant cell. The plasmid carries carbenicillin resistance (Deblaere *et al.*, 1985).

pJT62	A pUC12 based plasmid containing the CaMV 35S promoter and polyadenylation signal (Guerineau <i>et al.</i> , 1991). Obtained from Dr. Phil Mullineaux.
p486	A 4.2 kb <i>Pvu</i> II/ <i>Eco</i> RV genomic fragment cloned into pBluescript and containing the <i>CO</i> gene. Obtained from Kate Dixon.
p21	The <i>CO</i> cDNA amplified by PCR using primers designed to introduce <i>Hind</i> III/ <i>Bam</i> HI sites for cloning into pBluescript. Obtained from Dr. Rüdiger Simon.
p505	A genomic fragment containing the <i>CO</i> ORF under the control of the CaMV 35S promoter in the binary vector pSLJ1711. Obtained from Dr. Maribel Igeño.
pAP2	A 1 kb <i>Pst</i> I/ <i>Bam</i> HI clone containing part of the <i>APETALA 2</i> gene. Obtained from Dr. Renate Schmidt.
c04541	A T-DNA cosmid vector derived from the binary vector SLJ1711 (Jones <i>et al.</i> , 1992) by the insertion of a fragment containing a cos site between the <i>Bgl</i> II sites. Obtained from Dr. Clare Lister.
CIC9E2, CIC9F3, CIC8H6, CIC3F11, CIC8E12 and CIC8H5:	YACs in vector pYAC4 identified from the CIC library (Creusot <i>et al.</i> , 1995). Obtained from Dr. Melanie Stammers.

7.1.4 General Media and Solutions

L-broth	Difco bacto-tryptone 10g, yeast extract 5g, NaCl 10g per litre, pH 7.5
L-agar	As above with 15g agar per litre
YT-broth	Difco bacto-tryptone 8g, yeast extract 5g, NaCl 5g per litre, pH 7.4
LBglycerol	Difco bacto-tryptone 10g, yeast extract 5g, NaCl 5g, 500ml glycerol per litre pH 7.2

YEPD-agar	Difco bacto-peptone 20g, yeast extract 10g, glucose 20g, adenine 0.1g, 2% bacto agar per litre
YEP-broth	Difco bacto-peptone 10g, yeast extract 10g, NaCl 5g per litre
M9-agar	NaH ₂ PO ₄ 6g, KH ₂ PO ₄ 3g, NaCl 0.5g, NH ₄ Cl 1g pH 7.4, 1.5 % bacto agar per litre. Add 2ml 1M MgSO ₄ , 10ml 20% glucose and 0.1ml 1M CaCl ₂ after autoclaving.
20 x SSC	3M NaCl, 3.3M Na Citrate pH 7.0
20 x SSPE	3M NaCl, 0.2M NaH ₂ PO ₄ .H ₂ O, 20mM EDTA pH 7.4
TE pH 8.0	10mM Tris.Cl pH 8.0, 1mM EDTA pH 8.0
10 x TBE	0.9M Tris base, 0.9M Boric acid, 0.02M EDTA pH 8.0
10 x TAE	0.4M Tris-acetate, 10mM EDTA pH 8.0
10 x MOPS	0.2M 3-[N-Morpholino]-propanesulphonic acid pH 7.0, 50mM Na acetate, 10mM EDTA
10% SDS	100g Na dodecyl sulphate per litre water
SMθ buffer	KH ₂ PO ₄ 3g, Na ₂ HPO ₄ anhydrous 7g, NaCl 5g, 1ml (1M) MgSO ₄ , 1ml (0.1M) CaCl ₂ , 1% gelatin made up to 1 litre with water.
Loading dye	0.25% bromophenol blue, 40% glycerol in 1 x TBE

7.1.5 Plant Media

Vacuum infiltration Medium (VIM)

0.5 x Murashige and Skooge salts (MS), 1 x B5 vitamins, 5% sucrose. Add 0.044μM benzylamino purine and 0.03% Silwet L-77 (Union Carbide) after autoclaving

Germination medium (GM)

1 x Murashige and Skooge salts, 1% sucrose, 100μg/ml inositol, 1μg/ml thiamine, 0.5μg/ml pyridoxine, 0.5μg/ml nicotinic acid, 0.5mg/ml 2-(N-morpholino)ethanesulphonic acid (MES), pH 5.7 with 1M KOH, 0.8 % Difco bacto agar

GM K50

As GM but supplemented with 50µg/ml Kanamycin (Sigma)

GM K50 R20

As GM K50 but supplemented with 20µg/ml Rovral fungicide (Rhône-Poulenc) for selection of transformants after vacuum infiltration

Callus inducing medium (CIM)

1 x Gamborg's B5 medium (without 2,4-dichlorophenoxyacetic acid (2,4-D), kinetin and sucrose) (Flow Labs), 2% glucose, 0.5mg/ml MES pH 5.7, 0.8% agar, 0.5µg/ml 2,4-D (Sigma), 0.05µg/ml kinetin (Sigma)

Liquid IM

1 x Gamborg's B5 medium as above, 2% glucose, 0.5mg/ml MES pH 5.7

Shoot inducing medium (SIM V750 K50)

1 x Gamborg's B5 medium (as for CIM), 2% glucose, 0.5mg/ml MES pH 5.7, 0.8% agar, 5µg/ml N6-(2-isopentenyl)adenine (2ip) (Sigma), 0.15µg/ml indole-3-acetic acid (IAA) (Sigma), 750µg/ml vancomycin (Eli-Lilly), 50µg/ml kanamycin

7.2 Methods

7.2.1 Bacterial and General DNA Techniques

7.2.1.1 Methods of DNA Extraction

i) Alkaline lysis miniprep of plasmid DNA from *E. Coli* - based on the method of Birnboim and Doly (1979).

A single bacterial colony was inoculated into 10ml of L-broth with the appropriate antibiotic and grown overnight at 37°C with vigorous shaking. 1.5ml of this culture was pelleted for 30 seconds in a benchtop centrifuge. The pellet was resuspended by vortexing in 100µl ice cold solution I (50mM EDTA, 25mM Tris.Cl pH 8.0). 200µl freshly prepared solution II (0.2N NaOH, 1%SDS) was added and the mixture inverted several times before incubating at room temperature for 5 minutes. 150µl ice cold solution III (3M/5M potassium acetate, pH4.8) was added, the mixture vortexed briefly and incubated on ice for 5 minutes. The cell debris and chromosomal DNA was pelleted by centrifugation for 5 minutes and the supernatant was removed to a clean tube. This was extracted with phenol/chloroform then chloroform (optional), the DNA ethanol precipitated, dissolved in 20-50µl TE pH 8.0 or water and stored at -20°C. 1/10 of this solution was normally used for restriction enzyme digests.

ii) Large scale plasmid and cosmid DNA extraction

For large scale preparation of plasmid DNA a single bacterial colony was inoculated into 50-400ml L-broth or YT broth with appropriate antibiotics and grown overnight at 37°C with vigorous shaking. The bacterial cells were harvested by centrifugation at 5000rpm for 10 minutes in a Sorvall H6000A rotor. The plasmid DNA was extracted and purified using a midi/maxi plasmid qiagen kit (Qiagen) in accordance with the manufacturer's instructions.

iii) Small scale plasmid DNA extraction from *Agrobacterium* - based on the method of Koncz *et al.*, 1984.

10µl of a glycerol stock of the appropriate *Agrobacterium* strain was inoculated into 5ml YEP broth containing the appropriate antibiotic to select for the binary vector and Rifampicin (200µg/ml) to select for *Agrobacterium*. This culture was grown at 30°C for 48 hours with vigorous shaking. Cells from 4.5ml of culture were pelleted for 30 seconds and resuspended by gentle pipetting in 100µl solution I (25mM Tris.Cl pH 8.0, 10mM EDTA, 15% sucrose) with 4mg/ml freshly prepared lysozyme. It was then incubated for 5 minutes at room temperature followed by 5 minutes on ice. 200µl of freshly prepared solution II (0.1N NaOH, 1%SDS) was added and the solution gently mixed by inversion. Following another 5 minutes on ice 150µl ice cold solution III (3M/5M potassium acetate) was added, the mixture gently inverted and incubated on ice for 10 minutes. Cell debris and chromosomal DNA were pelleted by centrifugation for 10 minutes at 4°C and the supernatant extracted with phenol/chloroform then chloroform, ethanol precipitated, the pellet dissolved in 20µl TE pH 8.0 and stored at -20°C. 10µl of this solution was used in restriction enzyme digests which, because of the poor quality of plasmid DNA prepared from *Agrobacterium*, were usually Southern blotted and probed with an appropriate DNA fragment to visualize the bands of interest.

7.2.1.2 Glycerol Stocks

To preserve bacterial strains 0.75ml of overnight (*E. coli*) or 48 hour (*Agrobacterium*) cultures were mixed with 0.75ml of LB-glycerol in an Eppendorf tube, frozen briefly in liquid nitrogen and transferred to a -70°C freezer for storage.

7.2.1.3 Preparation of Competent *E. coli* Cells

To prepare cells for chemical transformation a single bacterial colony (usually *E. coli* strain DH5 α) from a freshly streaked plate was used to start a 10ml L-broth overnight culture. 1ml of this culture was used to inoculate 100ml of YT broth or L-broth which was grown at 37°C with shaking until the OD₆₀₀ reached 0.4-0.5 (approximately 2 hours). The cells were cooled on ice and harvested by centrifugation (3000rpm for 10 minutes at 4°C in a Sorvall H6000A rotor) followed by resuspension in 4ml ice cold TFB I buffer (30mM potassium acetate, 100mM RbCl₂, 10mM CaCl₂, 50mM MnCl₂ in 15% glycerol adjusted to pH 5.8 with 0.2M acetic acid and filter sterilized). The cells were again harvested by centrifugation, resuspended in 4ml ice cold TFB II buffer (10mM MOPS, 10mM RbCl₂, 75mM CaCl₂ in 15% glycerol adjusted to pH 6.5 with KOH and filter sterilized) and divided into 100 μ l aliquots. These were either kept on ice for immediate use or frozen briefly in liquid nitrogen and transferred to a -70°C freezer.

7.2.1.4 Transformation of *E. coli*

Frozen competent cells (100 μ l aliquots in 1.5ml Eppendorf tubes) were gently thawed on ice for at least 10 minutes. 1.7 μ l of β -mercaptoethanol (a fresh 1/10 dilution of a 14.4M stock) was added and the cells incubated on ice for 10 minutes with gentle flicking every 2 minutes to mix. 1-50ng of plasmid DNA was added to the cells, mixed by flicking and incubated on ice for 30 minutes. The cells were then heat-shocked by immersing the tubes in a 42°C water bath for 45 seconds before immediately being transferred back to ice for a further 2 minutes. 0.9ml of preheated YT broth was added and the mixture incubated at 37°C for 1 hour with vigorous shaking. Cells were plated on L-agar plates containing antibiotics to select for the transformed plasmid and where appropriate IPTG (50 μ g/ml) and X-Gal (40 μ g/ml) to select for recombinant plasmids in pBluescript-based vectors and grown

overnight at 37°C. Typically 0.1µl-100µl of the total transformation mixture (1ml) was plated.

7.2.1.5 Transfer of Plasmids from *E. coli* to *Agrobacterium tumefaciens*

i) Tri-parental mating - transfer of plasmids by conjugation based on Ditta *et al.* (1980)

Day 1; The *E. coli* "donor" strain containing the binary vector (usually pSLJ1711) and the *E. coli* "helper" strain HB101 containing the plasmid pRK2013 were streaked out on fresh selective plates containing 10µg/ml tetracycline and 25µg/ml kanamycin, respectively, and grown overnight at 37°C.

Day 2; A 24 hour culture using 10µl from a glycerol stock of *Agrobacterium tumefaciens* strain C58C1 was prepared in YT broth and grown at 30°C. Overnight cultures from single colonies of the two *E.coli* strains were also prepared. None of the cultures contained antibiotics.

Day 3; 1ml of the *Agrobacterium* culture and 0.2ml of each of the two *E. coli* cultures were transferred into one Eppendorf tube, mixed, and the cells pelleted by centrifugation for 1.5 minutes. The cell pellet was resuspended in 0.1ml of YT broth, spread onto an L-agar plate without antibiotics and allowed to grow at 30°C overnight.

Day 4; Approximately 3 loopfuls of the bacterial lawn were transferred into 2ml of 10mM magnesium sulphate and 100µl of 10⁻², 10⁻⁴ and 10⁻⁶ dilutions were plated on L-agar plates containing 1µg/ml tetracycline and 200µg/ml rifampicin. These were incubated at 30°C for 2-3 days until colonies appeared.

Day 6/7; Individual colonies were streaked out on M9-agar (minimal) plates containing the appropriate antibiotics and grown for 2-3 days to select for *Agrobacterium*.

ii) Direct transformation of *Agrobacterium* - based on Holsters *et al.* (1978)

A 10ml YEP broth culture inoculated with 10µl from a glycerol stock of *Agrobacterium tumefaciens* strain C58C1 and containing 200µg/ml rifampicin was grown overnight at 30°C

with shaking. 4ml of this culture was inoculated into 100ml fresh YEP broth plus 200µg/ml rifampicin and grown at 30°C for 4-5 hours until mid-log phase. The cells were pelleted by centrifugation at 2500rpm for 10 minutes in a Sorvall H6000A rotor before being resuspended in 2ml chilled YEP broth. This was divided into 200µl aliquots in prechilled Eppendorfs and chilled on ice for 5 minutes. DNA for each binary construct was prepared using a qiagen midi kit and diluted to 0.1-0.2µg/µl in water. 10µl (1-2µg) of DNA was pipetted into the cells and mixed by gently flicking the tubes. 10µl of water in one aliquot of cells served as a control. The aliquots of cells were frozen immediately in liquid nitrogen for 15 seconds and then heat-shocked in a 37°C water bath for 5 minutes. 1ml YEP broth was added to each tube and they were incubated at 30°C with shaking for 1-2 hours. 0.1ml aliquots were plated onto L-agar plus 1µg/ml tet, 200µg/ml rif. plates and incubated at 30°C for 1-5 days until colonies appeared. Individual colonies were streaked onto M9-agar plates with the appropriate antibiotics and grown at 30°C for 2-3 days until colonies appeared.

7.2.1.6 Preparation of Filters for Colony Hybridization

Colony hybridization was performed when a particular rare recombinant plasmid needed to be isolated from a large background of non-recombinant or unwanted recombinant plasmids, for example, after a difficult cloning step where the number of individual DNA minipreps required would be impractical or during the screening of a cosmid library. Cells were plated out on L-agar with appropriate antibiotics and grown at 37°C overnight. Details on the plating of cosmid libraries are given in the section "Preparation and screening of a Cosmid Library" (7.2.6). For isolation of a particular plasmid following a cloning step individual colonies were picked using a toothpick and plated into a grid pattern on L-agar plus appropriate antibiotics to make identification of the positively hybridizing colonies easier. Colonies expected to produce a positive or negative result were also included. After overnight incubation the colonies were transferred to nylon membrane (Hybond-N,

Amersham). A membrane of an appropriate size and dampened on a sterile L-agar plate, was placed on the surface of the plate, marked with 3 asymmetric pin holes and peeled off after 1 minute. Subsequent duplicate membranes were left on the surface for double the previous time. The pin holes were duplicated onto the underside of the plate. The membranes were handled and processed always with the colonies facing up. To fix the DNA from the colonies to the membrane they were treated sequentially with a number of solutions soaked into pads of Whatman's 3MM paper: 10% SDS for 5 minutes, 1.5M NaCl, 0.5M NaOH for 5 minutes, 1M Tris.Cl pH 8.0, 1.5M NaCl for 2 x 5 minutes and 2 x SSPE for 3 minutes, ensuring that the solutions did not come into direct contact with the upper surface. The membranes were then pressed between 2 sheets of clean 3MM paper to remove excess cell debris, air dried and baked at 80°C for 2-3 hours. A prewashing step (50mM Tris.Cl pH 8.0, 1M NaCl, 1mM EDTA, 0.1% SDS at 42°C for 1 hour) removed more cell debris before the membranes were prehybridized and hybridized as outlined in section 7.2.1.12.

7.2.1.7 Cleavage of DNA with Restriction Enzymes

All restriction enzyme digestions were carried out in 1.5ml Eppendorf tubes. Small scale digestions (<1µg) were typically carried out in small volumes of 20-30µl which fit conveniently into standard agarose gel wells. Large scale plasmid, cosmid, yeast or plant DNA digestions were carried out in much larger volumes, up to 300µl and typically contained 1mM spermidine (Sigma) and 33µg/ml BSA (Boehringer). If necessary the volume was subsequently reduced by ethanol precipitation. Heat-treated RNase A at 50µg/ml was included in DNA digestions where necessary. All digestions contained the appropriate restriction enzyme buffer supplied by the manufacturers at 1 x concentration plus 1-20 units of enzyme as appropriate. Plasmid DNA digests containing <1µg were incubated at the manufacturer's recommended temperature - usually 37°C - for 1-2 hours.

Plant and yeast genomic DNA digests and large scale plasmid or cosmid digests were incubated overnight. In double digests the two enzymes were added simultaneously if requiring the same buffer, or sequentially starting with the enzyme requiring the lowest salt concentration. All restriction enzymes were obtained from Pharmacia Ltd., BCL Boehringer, Northumbria Biologicals Ltd. or Gibco (BRL).

7.2.1.8 Agarose Gel Electrophoresis of DNA

Separation of DNA fragments according to size was performed by electrophoresis through horizontal submerged agarose gels. This was done routinely as a confirmatory technique during the construction of novel plasmids, to separate plant DNA prior to Southern blotting and to visualize the results of PCR. The gels used were typically 0.4-1.5% depending on the size of the fragments to be separated. All gels were made with and run in 1 x TBE buffer. Large gels were electrophoresed overnight at 35-70 volts whereas smaller gels were run at 50-120 volts for 1-3 hours. 1/10 volume of 10 x loading dye was added to each DNA sample prior to loading into preformed wells and in the case of plant DNA digests 1/10 volume of 1M NaCl was also added to even out the salt concentration between samples. Size markers were separated alongside the samples for comparison (*Hind* III digested lambda DNA or 1kb ladder (BRL)). To visualize the DNA 0.1µg/ml ethidium bromide was added to the cooled molten agarose before casting the gel and to the running buffer. Gels were examined on a short wave (610nm) UV transilluminator (UVP Inc.) and photographed onto Sony UPP-110HA thermal paper using a UVP ImageStore 5000 camera system attached to a Sony Video Graphic Printer.

7.2.1.9 Isolation and Recovery of DNA Fragments from Agarose Gels

i) Using the Qiaex and Qiaquick kits (Qiagen)

The restriction enzyme digestion from which the fragment was to be isolated was electrophoresed at 30 volts until the fragment of interest was clearly separated from all other fragments. If the fragment was to be used for cloning purposes it was visualized using a long-wave UV lamp (model UVSL.25, Mineralight) to minimise DNA damage. It was then excised from the gel using a clean scalpel blade and isolated from the gel according to the manufacturer's instructions.

ii) Using the Spin-X column kit (Costar)

The DNA fragment was excised from the gel as above, chopped finely and placed in the Spin-X column which was then frozen at -70°C for 1 hour. The column was allowed to thaw to room temperature before spinning at 4000rpm in a benchtop centrifuge for 30 minutes. The eluate was extracted once with phenol/chloroform followed by chloroform, ethanol precipitated and dissolved in a small volume of TE or water.

7.2.1.10 DNA Ligations

To ligate a fragment with 5' or 3' overhangs after restriction enzyme digestion ("sticky ends") into small, high-copy number plasmids (for example, pBluescript) typically 100ng linearized vector DNA and sufficient insert DNA to give a molar ratio of DNA "ends" of 1:3 (vector:insert) were used. For ligations into large binary vectors (for example, pSLJ1711 - 30kb) the ratio of vector to insert was the same but up to 1 μg of vector was used. All ligations were typically carried out in a volume of 10-30 μl in 1 x ligation buffer (50mM Tris.Cl pH 7.4, 10mM MgCl_2 , 10mM DTT, 1mM spermidine, 1mM ATP, 0.1mg/ml BSA) with 1-2 Weiss units of T4 DNA ligase (1u/ μl , BRL) at 15°C for 8 hours-overnight. Vector DNA digested with a single restriction enzyme was treated with calf intestinal alkaline phosphatase (CIAP - 0.5u/ μg DNA) in 1 x CIAP buffer supplied with the enzyme,

for 30 minutes at 37°C and the enzyme removed by phenol/chloroform extraction and ethanol precipitation prior to ligation. This treatment removes 5' phosphate groups from the DNA ends, preventing self-ligation. To ligate fragments with blunt ends the volume was reduced to 5-10µl and the ligation incubated overnight at 4°C in 1 x blunt end buffer (66mM Tris.Cl pH 7.4, 5mM MgCl₂, 5mM DTT, 0.5mM spermidine, 0.2mM ATP, 1mM hexamine cobalt chloride).

7.2.1.11 Southern Blotting

Following separation through agarose gels DNA was transferred to nylon membranes and hybridized to radioactively labelled DNA probes (Southern 1975, modified by Wahl *et al.* 1979). While this technique was used mainly for the detection of particular DNA sequences in plant DNA it was also routinely used for plasmid DNA (to monitor cloning steps), cosmid DNA (to aid restriction mapping) and for monitoring PCR and RT-PCR products. The gel was treated as follows: depurination of the DNA for 20 minutes in 0.25M HCl (except for fragments smaller than 10kb), denaturation for 30 minutes in 0.5M NaOH, 1.5M NaCl and neutralization for 2 x 15 minutes in 1M Tris.Cl pH 7.2, 1.5M NaCl. The gel was rinsed between each step with water and all steps involved several gel-volumes of solution with gentle shaking. The gel was moved to a piece of Whatman 3MM paper prewetted with 10 x SSC on a glass plate supported over a reservoir of 10 x SSC. The ends of the 3MM "wick" were in contact with the reservoir. A piece of Hybond-N (or Hybond-N+, Amersham) the same size as the gel was placed on top followed by 2 pieces of 3MM, prewetted in 10 x SSC. A stack of paper towels to draw moisture through the apparatus was pressed down on top with a glass plate and a 1kg weight (for a diagram see Maniatis *et al.* 1989). The blotting apparatus was generally left overnight although 2 hours was sometimes sufficient for plasmid DNA. The blot was dismantled, the membrane rinsed

briefly in 2 x SSC and allowed to air dry. The DNA was immobilized on the membrane by UV stratalinking at an energy setting of 2400 in a Stratagene "Stratalinker".

7.2.1.12 Hybridization of Radio-labelled DNA probes to DNA Immobilized on Nylon Membranes

This technique was used for the detection of particular DNA sequences from DNA transferred to membranes by Southern blotting (7.2.1.11) and by colony blotting (7.2.1.6). Membranes were prehybridized and hybridized either in tightly sealed polypropylene boxes in a moving-shelf hybridization oven (New Brunswick Scientific Co. Ltd.) or in pyrex tubes in a hybridization oven (Bachofer) with a rotating spindle. 25ml pre/hybridization solution was used for polypropylene boxes but in pyrex tubes after prehybridization in 25ml solution the volume during hybridization was reduced to 5-10ml. Colony hybridizations were usually performed in petri dishes contained within polypropylene boxes with sufficient pre/hybridization solution to cover the membranes. All hybridizations were carried out at 65°C unless otherwise stated. Several different methods were investigated for pre/hybridization and washing membranes but the most effective was based on the method of Church and Gilbert (1984). Membranes were prehybridized for 2-4 hours in 0.3M sodium phosphate buffer pH 7.2 (to make 1L: 1M Na₂HPO₄ 684ml, 1M NaH₂PO₄ 316ml), 7% SDS, 1mM EDTA, 1% w/v BSA. Denatured radio-labelled DNA probe was then added to the solution and hybridization allowed to proceed overnight. The hybridization solution was poured off into a suitable container (usually a 50ml tube) and stored at -20°C in a perspex box for re-use or to decay. The membranes were washed by first rinsing for 5 minutes in wash buffer (0.5M sodium phosphate pH 7.2, 5% SDS, 50mM EDTA) added at room temperature, then washed for 2 x 30 minutes in the same buffer preheated to 65°C in the pyrex tubes or boxes used for hybridization. Colony filters were removed from their petri dishes and washed in polypropylene boxes. The membranes and discarded wash

solutions were continually monitored using a Geiger counter and the number and/or duration of washes adjusted as appropriate. If the membranes were to be reused for subsequent DNA probes they were always kept moist in a sealed bag following washing. The labelled probe was removed by incubation in 1.5M NaCl, 0.5M NaOH at 45-55°C for 20-40 minutes with gentle shaking. The membranes were then neutralized using the same conditions in 1M Tris.Cl pH 7.2, 1.5M NaCl.

7.2.1.13 Autoradiography

While still moist after washing, the membranes were sealed in Saranwrap or an air-tight plastic envelope before exposure to Kodak X-OMAT AR film in cassettes with intensifying screens at -70°C. Exposure times ranged from 30 minutes (for plasmid DNA and PCR products) to 2 weeks (for plant DNA) depending on the strength of the radioactive signal. Films were developed in a Fuji X-Ray Film Developer RG II.

7.2.1.14 Radioactive Labelling of DNA Fragments

DNA fragments to be used as radioactive probes were radio-labelled by a method based on Feinberg and Vogelstein (1983). 20-50ng of gel-purified DNA fragment or 50-200ng of digested plasmid or cosmid DNA was denatured in a volume of 32µl for 5 minutes at 100°C and cooled on ice. The following reagents were added in this order: 5µl 10 x Hexanucleotide Mix (Boehringer Mannheim), 5µl 2.5mM dNTPs minus dCTP, 1µl 10mg/ml BSA, 1.85MBq [α -³²P]dCTP at a specific activity of 111 TBq/mmol and a concentration of 370 MBq/ml (NEN) and 1.5µl Klenow fragment of *E. coli* DNA polymerase I (1u/µl, NBL). The reagents were mixed and the reaction incubated at 37°C for 30 minutes-2 hours. To separate unincorporated nucleotides the reaction was passed over a sephadex G50 column equilibrated with TE after addition of 10µl "nucleotide dye" (1% dextran blue, 0.1% orange G in TE). The blue fraction containing the radioactively-labelled DNA was collected

and the slower orange fraction containing the unincorporated nucleotides was discarded. The approximate percentage of incorporation was estimated by comparing the radioactivity of the two coloured fractions using a Geiger counter held at a suitable distance from the collection tubes.

7.2.1.15 Quantification of Nucleic Acid Concentration

i) Spectrophotometric estimation

The optical density (OD_{260}) of a DNA or RNA sample was determined using a Philips PU 8620 UV/VIS/NIR spectrophotometer. The concentration was calculated where:

$$\begin{aligned} 1 \text{ } OD_{260} \text{ unit} &= 50\mu\text{g/ml double stranded DNA} \\ &= 40\mu\text{g/ml single stranded DNA/RNA} \\ &= 33\mu\text{g/ml oligonucleotide (see 7.2.3.6)} \end{aligned}$$

ii) UV quantification

An estimate of concentration was made by running the DNA sample on an agarose gel with several DNA standards (for example lambda DNA) of known concentration. The intensity of each band visualised on a UV transilluminator after staining the gel with ethidium bromide was compared to the known standards.

7.2.1.16 Ethanol Precipitation of DNA

A 1/10 volume of 3M sodium acetate pH 5.2 was added to the DNA solution and mixed well. Two volumes of ice-cold ethanol were added, mixed well and stored at -20°C for 30 minutes-overnight. The DNA was pelleted by centrifugation in a benchtop centrifuge at full speed or an SS34 rotor at 15,000rpm for 10-30 minutes. The pellet was washed with 70% ethanol (1ml in an Eppendorf, 5ml in 30ml Corex tubes, ensuring all walls of the tube were rinsed) and respun for 2 minutes if necessary, to re-attach the pellet to the bottom of the tube. The washing step was repeated if a lot of salt was present in the sample or if the

DNA pellet was large. The pellet was drained and dried at room temperature or briefly at 37°C or in a vacuum dessicator. The pellet was dissolved in a suitable volume of sterile distilled water or TE.

7.2.1.17 Isopropanol Precipitation of DNA

To precipitate small amounts of DNA and to minimise salt precipitation 1 volume of room temperature isopropanol (Propan-2-ol) was used in place of 2 volumes of ethanol. Unless otherwise stated the DNA was pelleted immediately after addition of the isopropanol.

7.2.1.18 Phenol/Chloroform Extraction of DNA

This was used to remove proteins from DNA preparations (after plant/bacterial DNA extraction or treatment with restriction or modifying enzymes). An equal volume of a 25:24:1 mixture of phenol:chloroform:iso-amyl alcohol equilibrated with TE pH 8.0 was added to the DNA solution. The aqueous and organic phases were mixed by inversion (for large fragments of DNA) or vortexing (for small fragments). The tubes were centrifuged for 5 minutes in a suitable centrifuge and the upper (aqueous) phase removed to a clean tube without disturbing the interphase. If necessary the extraction was repeated before a final extraction with an equal volume of 24:1 chloroform:iso-amyl alcohol to remove traces of phenol.

7.2.2 Plant DNA Extraction Methods

7.2.2.1 CTAB (Cetyl trimethyl ammonium bromide) Plant DNA Extraction

A modified miniprep CTAB protocol was used to prepare plant DNA for Southern analysis and PCR (Dean *et al.* 1992). 2-5 grams of frozen leaf material was ground to a fine powder using a precooled pestle and mortar. The powder was transferred to a 50ml polypropylene

tube (Corning), 25ml extraction buffer added (140mM Sorbitol, 220mM Tris.Cl pH 8.0, 22mM EDTA, 800mM NaCl, 1% N-Lauryl sarcosine, 0.8% CTAB) and the mixture incubated for 20 minutes at 65°C with occasional vigorous shaking. After addition of 10ml chloroform and inverting for 20 minutes at room temperature the samples were centrifuged at 2500rpm for 15 minutes in an H6000A rotor to resolve the phases. The nucleic acid was precipitated from the aqueous phase by the addition of 17ml of isopropanol, placed on ice for 10 minutes and centrifuged as before. The pellet was drained and dissolved in 4ml TE pH 8.0 and an equal volume of 4M lithium acetate added to precipitate RNA. The mixture was placed on ice for 20 minutes and after centrifugation (as before) the supernatant was removed to a clean tube and the DNA precipitated with 2 volumes of ethanol. Following an incubation on ice for 20 minutes and centrifugation the pellet was again dissolved in 4ml TE, extracted with phenol/chloroform and precipitated with 1/10 volume 3M sodium acetate and 2 volumes of ethanol. The resulting pellet was dissolved in 100-200 μ l TE. The yield was usually >10 μ g g^{-1} tissue.

7.2.2.2 Dellaporta DNA Microprep

This protocol was used for small scale DNA preparations. Plant material (up to 0.15g) was collected in Eppendorf tubes, immediately frozen in liquid nitrogen and transferred to a -70°C freezer until required. To extract the DNA the tube was cooled in liquid nitrogen and the tissue ground to a fine powder using a precooled microfuge pestle (Kontes) attached to a drill. 0.7ml extraction buffer (50mM Tris.Cl pH 8.0, 10mM EDTA, 100mM NaCl, 1% SDS, 1mM β -mercaptoethanol) was added, the tube vortexed to resuspend the pulverised tissue and incubated at 65°C for 10 minutes. 200 μ l 5M potassium acetate was added, the tube vortexed and incubated on ice for 20 minutes before being spun in a benchtop centrifuge for 10 minutes. The supernatant was transferred to a new tube and if it still contained insoluble material, respun and transferred as before, and the DNA precipitated

with an equal volume of isopropanol. The tube was spun immediately for 2 minutes and the DNA pellet washed with 80% ethanol. The pellet was dried at room temperature and dissolved in 50µl TE. If the DNA was to be used for PCR it was passed through a Sepharose CL-6B spin column. Restriction digests for Southern blotting contained 1/10 final volume 40mM spermidine. This method usually yielded sufficient DNA for 2 lanes on a Southern blot from 0.15g tissue.

7.2.3 Polymerase Chain Reaction Techniques

The polymerase chain reaction (PCR) was used to amplify genomic sequences for cloning, sequencing and as a diagnostic tool in screening as well as for the amplification of specific mRNA species following reverse transcription (RT-PCR). Protocols for RT-PCR are described in section 7.2.6 (RNA techniques).

7.2.3.1 PCR Amplification of Genomic and Plasmid DNA

Genomic and plasmid DNA was amplified essentially as described by Saiki *et al.*, 1988. 2ng DNA was used in each reaction, with 0.5µl *Taq* polymerase (Boehringer Mannheim), 0.25mM nucleotides (dATP, dCTP, dGTP, dTTP), 1µM primers in 1 x PCR buffer (10mM Tris.Cl pH 8.3, 50mM KCl, 1.5mM MgCl₂, 0.1mg/ml gelatin, 0.05% Tween 20). Amplification usually involved 35 cycles, each consisting of 1 minute at 94°C (denaturation), 1 minute at a temperature chosen usually between 50 and 65°C (annealing - see below), and 1.5 minutes at 72°C (extension). The amplification cycles were preceded by 4 minutes at 94°C to denature the DNA during which time the enzyme in 1/5 reaction volume in 1 x PCR buffer was added to prevent non-specific amplification before cycling started. The cycles were followed by 10 minutes at 72°C to complete synthesis. PCR was performed in a total volume of 50µl in a 0.5 ml Eppendorf tube. Mineral oil was added (20µl) if the PCR machine in use did not have a top heating block. Several PCR machines

were used, most frequently the Programmable Thermal Controller (M J Research Inc.) and the Cyclogene (Techne). Depending on the DNA to be amplified, the length of the expected product and the composition of the primers, several parameters were often changed: most frequently, the amount of DNA in each reaction, the extension time, the number of cycles and the annealing temperature. PCR products were analyzed by running 1/10 of the reaction on a gel, followed by Southern blotting and hybridization to radioactively-labelled DNA probes as appropriate.

7.2.3.2 Rapid Preparation of Plant Material for Diagnostic Screening by PCR

This technique was used when a large number of plants was to be screened by PCR as it bypasses the DNA extraction step. The top of an Eppendorf tube was used to sever a sample of leaf material (2mm²) into the tube containing 40µl of 0.25M NaOH. The submerged leaf was placed in a boiling water bath for 30 seconds, neutralized by the addition of 40µl 0.25M HCl followed by 20µl 0.5M Tris.Cl pH 8.0 and again placed in the boiling water bath for a further 2 minutes. The extract was then used as a PCR template or stored at 4°C. Stored samples were denatured for 2 minutes in a boiling water bath immediately before use (Klimyuk *et al.*, 1993).

7.2.3.3 Inverse Polymerase Chain Reaction (IPCR)

This was used to isolate unknown DNA adjacent to the deletion in *co-8*. 2µg of CTAB-extracted DNA was digested overnight in a volume of 300µl at 60°C with *Bst*Y I. The digest was ethanol precipitated and the pellet dissolved in 40µl water. 20µl (1µg) was self-ligated in a total volume of 300µl in 1 x ligation buffer (see 7.2.1.10) with 1µl ligase (BRL) at 15°C overnight. The ligation reaction was then extracted with phenol/chloroform and chloroform, ethanol precipitated and dissolved in 10µl water. 2.5µl was used as a PCR template in 50µl reactions (see 7.2.3.1) using primers adjacent to the *co-8* deletion and to

the known *Bst*Y I site. 1/10 of the reaction was run on a 1% agarose gel. Where a band of a predictable size was obtained the remainder of the PCR reaction was treated as described in section 7.2.3.4 (Cloning PCR products).

7.2.3.4 Cloning PCR Products

Following the examination of 1/10 of the PCR reaction on an agarose gel the remainder was phenol/chloroform and chloroform extracted, ethanol precipitated and the pellet dissolved in 20µl TE. To blunt-end the PCR product in preparation for cloning into the *Eco*R V site of pBluescript, 19.5µl of the PCR reaction was incubated in 1 x T4 DNA polymerase buffer (BRL), 6µl 2.5mM dNTPs, 1µl T4 DNA polymerase (5u/µl, BRL) and 0.5µl 10mg/ml BSA in a total volume of 30µl for 30 minutes at 37°C. The T4 DNA polymerase was denatured by incubation at 70°C for 10 minutes and the reaction fractionated on an agarose gel. The band of interest was excised using a clean scalpel blade and extracted from the agarose using one of the techniques outlined in section 7.2.1.9. Subsequent ligations, usually into *Eco*R V cut pBluescript, and transformation into *E. coli* strain DH5α followed the methods described in sections 7.2.1.4 and 7.2.1.10.

7.2.3.5 Colony PCR

This technique was used to amplify cloned DNA inserts directly from a large number of transformed *E. coli* colonies as a less time-consuming alternative to plasmid DNA extraction. The required number of wells in a microtitre dish were filled with 50µl LB plus the appropriate antibiotic. Single *E. coli* colonies were touched with the thin end of sterile toothpicks and placed into individual wells. 50µl aliquots of water were pipetted into the same number of 0.5ml eppendorf tubes and the toothpicks were transferred from the microtitre dish into the tubes for a few seconds and then discarded. The microtitre dishes were closed, wrapped in clingfilm and transferred into a 37°C incubator for the rest of the

day to provide cultures for glycerol stocks when the correct clones were identified. The 0.5ml tubes were closed and dipped into liquid nitrogen for 2 minutes immediately followed by 10 minutes in a boiling water bath and a brief spin in a benchtop centrifuge. A mix containing the following reagents was prepared; Per reaction: 100ng of each of the pair of primers (usually M-13 forward and reverse), 10 μ l 10 x PCR buffer (see 7.2.3.1), 0.5 μ l *Taq* polymerase (Boehringer Mannheim), 2.5 μ l 4mM dNTPs and water to 50 μ l. 50 μ l of the mix was added to each of the PCR tubes plus 20 μ l mineral oil. Amplification was as described in section 7.2.3.1.

7.2.3.6 Oligonucleotide Preparation

Oligonucleotides were synthesised using a Pharmacia LKB Gene Assembler Plus. To detach the oligonucleotide from the support column following synthesis, the support was incubated in a screw-capped, sealed tube in concentrated ammonia solution (35%) at 55°C overnight. After cooling on ice for 15 minutes the whole apparatus was spun in a benchtop centrifuge for 1 minute, the support removed to a clean tube and respun, and the extracted ammonia pooled. The oligonucleotide was precipitated with ethanol, spun for 30 minutes, washed and dissolved in 1ml TE pH 8.0. After measuring the absorbance at 260nm (see 7.2.1.15) of a 5 μ l sample diluted to 1ml, the concentration of the sample was calculated using the following formula: Concentration=Optical Density/Extinction co-efficient. To determine the extinction co-efficient (ϵ), the number of A,C,G and T nucleotides was counted. ϵ was then calculated as: $A(15.4 \times 10^3) + G(13.7 \times 10^3) + C(9.1 \times 10^3) + T(7.4 \times 10^3)$. The following oligonucleotides were used in PCR and sequencing experiments:

NAME	COMPOSITION
Bnc07	5'- GAATTCCGCCATAAAAGGGTCAGA - 3'
Bnc09	5'- AAGCTTTTGATGGATTGTAT-GCGT - 3'

Oli9	5'- AAATGTATGC-GTTATGGTTAATGG - 3'
Oli5	5'- CATTAACCATAAC-GCATACATTTC - 3'
AP2 Oli3	5'- CTCAATGCCG-AGTCATCAGG - 3'
AP2 Oli4	5'- CATGAGAGGAGGTTGGAAGC - 3'
G775	5'- GACTCGAGTCGACATCGAT ₁₇ - 3'
CO17	5'- ATGGATCATGTGGACTAG - 3'
CO19	5'- CATGGCAATACACGGTGC - 3'
CO20	5'- GAGAGTAACGACATAGG - 3'
CO21	5'- GATTCGAGCTAGCACGTG - 3'
CO25	5'- TACTGTTGTGCAAATGG - 3'
CO28	5'- TGCAGATTCTGCCTACTTGTGC - 3'
CO32	5'- TGAGAGAAGTTCCAAACCGC - 3'
CO33	5'- GAACAACAGGCACGACCC - 3'
CO34	5'- GACAATAAGGTATGCTTCG - 3'
CO40	5'- CACAGACGAGCTACGGGGG - 3'
CO41	5'- GGTCCCAACGAAGAAGTGC - 3'
CO42	5'- CAGGGAGGCGTGAAAGTGT - 3'
CO43	5'- GGGGATCTGTACCAATAGACTG - 3'
CO48	5'- CACCTTCCTCTTGATCCAC - 3'
CO49	5'- GCTCCACACCATCAAAC TTACTAC - 3'
CO50	5'- CTCCTCGGCTTCGATTTCTC - 3'
CO51	5'- CCATTGTCGTTGTAGTGAGTCC - 3'
CO52	5'- GGAACAGCCACGAAGCAAC - 3'
CO53	5'- ACGCCATCAGCGAGTTCC - 3'
CO54	5'- ACCCAAGAACGCCAAAGG - 3'
CO55	5'- CAGGGTCAGGTTGTTGCTCT - 3'

CO56	5'- CCCCCGTAGCTCGTCTGTG - 3'
CO57	5'- AGCATCTTCCCATAACGCAG - 3'
CO58	5'- TCGAAGCATACCTTATTG - 3'
CO65	5'- GCGTCACTCGCTTTGGTGGTGAGTAGTGG - 3'
CO74	5'- GATGGGCTACGTATGCGGC - 3'
CO75	5'- GGACTAAGCATATACGACACATCTC - 3'
M13-N	5'- GTAAAACGACGGCCAGT - 3'
M13-R	5'- AACAGCTATGACCATG - 3'
T3	5'- ATTAACCCTCACTAAAG - 3'
T7	5'- AATACGACTCACTATAG - 3'

Hyphens interrupting primer sequences oli5, oli9, AP2 Oli3, and bnco9, represent the position of introns where primers were designed to specifically amplify cDNAs.

7.2.4 RNA Extraction and Manipulation

7.2.4.1 Small Scale Total RNA Extraction from Plant Tissue

The following technique, based on the method of Stiekema *et al.*, 1988, was used to extract RNA from *Arabidopsis* seedlings for Northern blotting (see 7.2.4.2) and RT-PCR (see 7.2.4.3). 10-20 seedlings, 8-14 days old, were harvested into an Eppendorf, frozen in liquid nitrogen and transferred into a -70°C freezer until required. To extract RNA each tube was first filled with liquid nitrogen. This was allowed to evaporate off before grinding the tissue to a fine powder using a precooled microfuge pestle (Kontes) attached to a drill. 500µl of extraction buffer (80mM Tris.Cl pH 9.0, 150mM LiCl, 5mM EDTA, 5% SDS) was added and the contents homogenized before placing on ice. The contents were extracted twice with phenol/chloroform and once with chloroform. 1/3 volume of 8M LiCl was added to a final concentration of 2M and the RNA allowed to precipitate at 4°C overnight. The RNA

was pelleted by centrifugation for 10 minutes and the pellet washed with cold 85% ethanol before drying briefly at room temperature. If the RNA was to be used for Northern analysis only (not RT-PCR) a quick protocol was followed. The pellet was dissolved in 200 μ l DEPC-treated water, precipitated with 1 volume of isopropanol and 1/10 volume of 3M sodium acetate for 1 hour at -20°C. The RNA was pelleted by centrifugation for 10 minutes, washed and dried as before and dissolved in DEPC-treated water to a final concentration of 5 μ g/ μ l. If the RNA was to be used for RT-PCR, after the overnight LiCl precipitation the RNA was pelleted by centrifugation for 10 minutes and dissolved in 100 μ l 0.5 x React 3 restriction enzyme buffer (Gibco BRL) containing 1 μ l DNase I (Gibco BRL) and 1 μ l RNase inhibitor (RNAguard, Pharmacia). This was incubated at 37°C for 30 minutes-2 hours to degrade contaminating DNA, before being extracted with phenol/chloroform as before. The final supernatant was precipitated with 1/10 volume 3M sodium acetate and 1 volume of isopropanol at -20°C for 1 hour, spun for 10 minutes at 4°C and the pellet washed with cold 85% ethanol. The pellet was dried briefly and resuspended in DEPC-treated water to a final concentration of 0.5-1 μ g/ μ l.

7.2.4.2 Northern Analysis

Based on the method of Goldberg (1980). All plastic-ware (the gel tank, gel former and comb) was soaked in 0.1M NaOH, 1mM EDTA for several hours prior to use, then rinsed in DEPC-treated water. RNA was fractionated by electrophoresis through 1.3% agarose, 2% formaldehyde (1/20 final gel volume of a 40% solution, added after the dissolved agarose had cooled to 50°C), in 1 x MOPS buffer (see 7.1.4). Samples contained 5-20 μ g total RNA in 50% de-ionized formamide, 7% formaldehyde, 0.5mg/ml ethidium bromide, 1mM EDTA, 20% glycerol, 1 x MOPS buffer, 0.25% bromophenol blue, and were heated to 65°C for 15 minutes before loading onto the gel. RNA ladder (0.16-1.77kb, Gibco BRL) was loaded alongside the samples. Electrophoresis was in 1 x MOPS buffer at 30-60 volts for

2-4 hours in a fume hood. After electrophoresis the RNA was visualized on a short-wave UV transilluminator and photographed as previously described for DNA electrophoresis. The RNA was transferred to Hybond-N membrane as described in section 7.2.1.11 (Southern blotting) except that the gel required no pretreatment before setting up the capillary blot. The RNA was bound to the membrane by UV cross-linking. The conditions for DNA-RNA hybridizations and subsequent washings were identical to those described for Southern blotting (see 7.2.1.12, Church and Gilbert, 1984). Following autoradiography RNA membranes were stripped by incubation in boiling 0.1% SDS poured onto the membrane and allowed to cool to room temperature.

7.2.4.3 RT-PCR (Reverse transcription - Polymerase Chain Reaction)

Based on the method of Frohman, Dush and Martin (1988).

A cocktail containing the following reagents was prepared:

2µg RNA (heated to 80°C for 2 minutes and cooled on ice)

2µl RNAsin (RNAguard, Pharmacia)

2µl 0.1M DTT

1µl 20mM dNTPs

1µl (1µg) Primer G775 (see 7.2.3.6)

1µl M-MLV Reverse Transcriptase (Gibco-BRL)

4µl 5 x Reverse Transcriptase buffer (Gibco-BRL)

Sterile distilled water to 20µl

The reaction was incubated at 37°C for 2 hours then diluted 10-fold with water. This stock could then be stored at -20°C providing sufficient template for 40 PCR reactions. PCR was carried out as described in section 7.2.3.1, using 5µl of the reverse transcription reaction in a final volume of 50µl. Increasing the concentration of MgCl₂ from 1.5mM to 2.5mM gave more reliable amplification. Gene-specific primers were used with one of the pair

designed to span an intron to rule out the possibility of amplification of contaminating DNA. The number of cycles for each pair of primers was adjusted so that when 1/5 of the reaction was analysed by agarose gel electrophoresis the amplified band was only just visible to the naked eye in control samples. At this point the PCR reaction would not have reached saturation and the amount of product was still increasing exponentially and would give an indication of the relative amount of mRNA in each sample. After fractionation the DNA was analyzed by Southern hybridization

7.2.5 DNA Sequencing

7.2.5.1 Sequencing Reactions

Double-stranded DNA templates of recombinant plasmid DNA, prepared using qiagen kits (see 7.2.1.1), were used in all sequencing reactions. Templates were sequenced by the dideoxy chain termination method described by Sanger *et al.*, 1977. The protocol used was modified after Chen and Seeburg (1985). Usually templates were processed in batches of 10. For one reaction 18µl of plasmid DNA (0.5µg/µl) was mixed with 5µl 1M NaOH, 1mM EDTA and incubated at 37°C for 15 minutes to denature the DNA. It was then purified by spin dialysis through a 200µl column of Sepharose CL-6B (Pharmacia) equilibrated with TE. The DNA was sequenced using Sequenase kits (USB) essentially as described in the manufacturer's protocol booklet. 7µl of eluate from the spin column was added to 2µl sequenase buffer (200mM Tris.Cl pH 7.5, 100mM MgCl₂, 250mM NaCl) and 1µl (10ng) primer and annealed at 37°C for 30 minutes. DNA was first labelled by primer extension in the presence of limiting amounts of dNTPs. A mix of the following reagents was prepared, sufficient for the number of reactions plus one: Per reaction: 1µl DTT, 2µl labelling mix (1.5µM dCTP, dGTP, dTTP), 0.5µl α-³⁵S dATP (5µCi, 10µM), 3 units sequenase (2µl). 5.5µl of the mix was added to the template and left at room temperature

for 2-5 minutes. A chain termination step followed where 3.5 μ l of the labelling reaction was added to each of 4 prewarmed tubes containing 2.5 μ l of either G, A, T or C termination mix (80 μ M for each dNTP, 50mM NaCl, 8 μ M for the appropriate ddNTP) and incubated at 37°C for 2-3 minutes. Reactions were terminated by the addition of 4 μ l of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) and stored at -20°C (for maximum of 1 week) or used immediately.

7.2.5.2 Polyacrylamide Gel Electrophoresis

DNA in sequencing reactions was denatured by boiling for 2 minutes, transferred to ice and loaded onto 6% polyacrylamide gels in 1 x TBE buffer (high pH 10 x TBE for sequencing: 1.3M Tris-base, 0.44M boric acid, 0.025M EDTA). To make the gels 70ml 6% (w/v) Sequencing Gel Solution (Anachem) was polymerized with 70 μ l 25% ammonium persulphate and 70 μ l TEMED (Bio-Rad). This is sufficient for 1 wedge gel cast between two 20cm x 50cm glass plates separated by wedge spacers (IBI Kodak Ltd.). The electrophoresis was carried out at 35 Watts in 1 x TBE and continued until the bromophenol blue dye-front had migrated just off the bottom of the gel. For longer runs electrophoresis was continued for a further 3-5 hours. The gels were fixed in 10% methanol, 10% acetic acid for 40 minutes, mounted on Whatman 3MM paper and dried under vacuum at 80°C for at least 1 hour. Gels were exposed to X- Omat AR Kodak X-ray film at room temperature.

7.2.5.3 Analysis of DNA Sequences

All sequence analysis was performed using the programs available in the Wisconsin Package: Genetics Computer Group (GCG) sequence analysis software package, version 7.2 (1992).

7.2.6 The Preparation and Screening of a Cosmid Library

Plant DNA from the *Arabidopsis thaliana* *co-8* mutant was extracted using the CTAB DNA extraction procedure (7.2.2.1) and further purified on a caesium chloride gradient (details in Maniatis *et al*, 1989).

i) Small scale test partial digestion; 10µg genomic DNA was added to a mix on ice containing 15µl 10 x A buffer (Boehringer Mannheim), 4mM spermidine, 1 x BSA to a total volume of 150µl which was then aliquotted into 9 tubes (first tube = 30µl, rest = 15µl). *Sau* 3A restriction enzyme was diluted in 1 x A buffer to 1u/µl and 1 unit added to the first tube. A series of 7 two-fold dilutions of the enzyme in the DNA mix followed with the ninth aliquot serving as an undigested control. All aliquots were incubated at 37°C for 1 hour, digestion stopped by the addition of 1µl 0.5M EDTA on ice and the aliquots run on a 0.4% agarose gel. Lambda DNA digested with Xho I and Xba I was the size marker used. The sample containing the most DNA in the size range 15-25kb was estimated. Half this amount of enzyme should give the maximum amount of molecules in this range.

ii) Large scale digest; Starting with 150µg genomic DNA a large scale digest was set up using the conditions determined during the test partials and 1µg removed for analysis on an agarose gel. The rest of the reaction was stopped by incubation at 65°C for 20 minutes followed by phenol/chloroform extraction and ethanol precipitation.

iii) CIP treatment of the DNA; The DNA was resuspended in 270µl water, 30µl 10 x CIP buffer and 2.5u CIP (Calf intestinal alkaline phosphatase) added and the mixture incubated at 37°C for 20 minutes. The enzyme was heat-inactivated at 68°C for 15 minutes followed by phenol extraction and ethanol precipitation. The DNA pellet was dissolved at a concentration of 75µg/300µl.

iv) Sucrose density gradient centrifugation; To size fractionate the DNA sucrose gradients were prepared according to Maniatis using 10% and 40% sucrose containing 1M NaCl,

20mM Tris.Cl pH 8.0, 5mM EDTA, mixed in a gradient mixer and aliquotted into 13.2ml cellulose centrifuge tubes (Kontron Instruments) using a peristaltic pump. Each 75µg sample was heated to 65°C for 10 minutes and loaded on top of a gradient. Gradients were balanced using mineral oil. Centrifugation was at 38,000rpm overnight at 20°C in a Sorvall Swing-out 41 rotor in a Sorvall OTD55B ultracentrifuge (DuPont). 300-400µl fractions were collected from a pin-hole in the bottom of the gradients and 10µl of every third fraction analysed by agarose gel electrophoresis as before. The fractions containing DNA in the size range 15-25kb were pooled and dialysed overnight against TE. The volume was reduced to 400µl by butanol reduction, the DNA ethanol precipitated and dissolved in 20µl water. The approximate concentration was determined by agarose gel electrophoresis of a 0.5µl sample against known concentration standards.

v) Preparation of vector DNA; 10µg caesium-pure 04541 vector DNA (see 7.1.3) was digested with *Bam* HI, an aliquot checked on a gel and the remainder phenol/chloroform extracted, ethanol precipitated and resuspended at 250ng/µl in water.

vi) Ligation of vector and insert; 1µg purified insert DNA was added to 0.25µg vector in a volume of 7.5µl. 0.5µl was removed as an unligated control. The remainder was incubated at 65°C for 5 minutes, 42°C for 20 minutes and room temperature for 2 hours to promote annealing of compatible ends. 1µl of 0.5M Tris.Cl pH 8.0, 100mM MgCl₂ was added along with 0.5µl 0.1M ATP, 0.5µl of a 5/1000 dilution of β-mercaptoethanol and 1u of T4 DNA ligase. The reaction was incubated at 12°C for 18 hours. The unligated control and 0.5µl of ligation were checked on a gel to determine the success of the ligation, visible as a decrease in mobility compared to the unligated control.

vii) Preparation of plating cells; The *E. coli* strain XL1 Blue MR (see 7.1.2) was streaked out on an L-agar plate and grown at 37°C overnight. A single colony was used to inoculate 50ml LB-Maltose (L-broth supplemented with 0.4% maltose) plus 0.5ml 1M MgSO₄. This was grown at 37°C with shaking for 4-6 hours until the OD₆₀₀ < 1.0. The bacteria were

pelleted at 2000rpm for 10 minutes, resuspended in 20mM MgSO₄ and the OD₆₀₀ adjusted to 0.5.

viii) Packaging; Gigapack II Gold packaging extracts (Stratagene) were used according to the manufacturer's instructions to package the recombinant cosmids. 1µl from the 10µl ligation was packaged in a pilot experiment and resuspended in 500µl SMθ buffer (see 7.1.4) with 20µl chloroform.

ix) Titering the test packaging reaction; 10µl from the 500µl packaging reaction was added to 100µl SMθ buffer and 200µl plating cells and incubated at 37°C for 20 minutes. 1ml LB was added and the reaction incubated for a further 45 minutes. 0.1ml and 0.5ml aliquots were spread onto L-agar tet₁₀ plates plus IPTG and X-Gal (see 7.2.1.4) and incubated at 37°C overnight. The number of colonies was counted and the titre of the library determined. The remainder of the ligation was divided into 2µl aliquots, packaged and titered as above.

x) Amplification of the cosmid library; The packaged aliquots in SMθ buffer were pooled, mixed with a equal volume of freshly prepared plating cells and incubated in a 15ml Falcon tube for 30 minutes at room temperature. 4 volumes of LB were added and the tube incubated at 37°C for 1 hour with shaking. The culture was spun at room temperature for 10 minutes, the pellet resuspended in 500µl LB and spread onto a 150mm L-agar tet₁₀ plate. After overnight incubation at 37°C the colonies were gently resuspended in 2 washes each of 3ml LB using a glass spreader. The LB was pooled into a 50ml tube and sterile glycerol added to a final concentration of 18% and tetracycline to 10µg/ml. The tube was gently rocked overnight at 4°C, divided into 1ml aliquots and frozen at -70°C. Using 2µl of the amplified library serial dilutions in LB from 10⁻²-10⁻⁹ were plated as before and the titre determined. This was always rechecked prior to each library screen.

xi) Screening the cosmid library; Roughly 50,000 colonies divided between 10 x 150mm L-agar tet₁₀ plates were used for each screen. The colonies were transferred to nylon membranes and processed as described in section 7.2.1.6 and screened using radioactively-

labelled DNA probes as described in section 7.2.1.12. Following autoradiography, "positive" colonies were identified and a plug of agar plus usually several colonies from the correct region was cored out of the plate using the wide end of a sterile pasteur pipette and resuspended in 1ml LB. This was in preparation for the isolation of single positive colonies. The titre was determined using serial dilutions as previously described and the colonies transferred to nylon membranes for re-hybridization (2nd round). After identification of positive colonies they were subjected to a 3rd round of screening to ensure their purity. DNA was isolated from the cosmids of interest as described in section 7.2.1.1(ii).

7.2.7 Yeast Techniques

7.2.7.1 Screening the CIC YAC (Yeast Artificial Chromosome) Library

The CIC YAC library (Creusot *et al.* 1995) was prepared from *Arabidopsis thaliana* Columbia DNA in vector pYAC4 and contains 1152 clones maintained in the yeast strain AB1380. The clones are maintained in a numbered grid as glycerol stocks in twelve 12 x 8 microtitre plates and on YEPD plates to provide a stock of cells for colony hybridisation and preparation of yeast DNA from individual clones of interest. The entire library can be represented twice on one colony hybridization filter which is designed so that each clone can be identified from its two unique positions on the grid. The CIC YAC library colony hybridization filters (Hybond-N+) were kindly provided by Dr. Melanie Stammers. The filters were prehybridized for >4 hours in 20ml PEG buffer (7% SDS, 0.25M Sodium phosphate (500ml 0.5M phosphate buffer per litre PEG buffer), 0.25M NaCl, 1mM EDTA, 0.02M PEG 6000) at 62°C. The radioactively-labelled DNA probe was added and the filters hybridized overnight in the same buffer at 62°C. The filters were washed by first rinsing them in 3 x SSC, 0.1% SDS briefly at room temperature. They were then washed for 10 minutes at 62°C in preheated 3 x SSC, 0.1% SDS then for 10 minutes at the same

temperature in 0.1 x SSC, 0.1% SDS. The DNA to be used as a radioactive probe was gel purified away from vector sequences twice to avoid hybridization of contamination vector sequences to the pYAC4 vector.

7.2.7.2 Yeast Genomic DNA Extraction

To extract DNA from yeast, the colonies of interest identified from the CIC YAC library screen were streaked out into 2cm squares from the YEPD stock plates onto fresh YEPD plates. The colonies were grown at 30°C for 2 days. As many cells as possible were scraped onto the end of a toothpick and resuspended in 400µl TE/SDS (0.4% SDS, 75mM Tris.Cl pH 7.5, 40mM EDTA) by vortexing. 400µl phenol/chloroform was added, mixed by vortexing and the tube transferred to 65°C for 15-30 minutes. After the incubation the contents of the tube were again vortexed and spun in a microcentrifuge for 10 minutes. The supernatant was transferred to a clean tube and ethanol precipitated. The precipitate was pelleted by centrifugation for 10 minutes and dissolved in 50µl water. After incubation at 4°C for 3-4 hours residual SDS was spun out for 10 minutes, the supernatant re-extracted with phenol/chloroform, ethanol precipitated and the DNA pellet dissolved in 20µl water. To analyse the YAC DNA each yeast genomic DNA preparation was digested with *Bam*HI/*Eco*RI in a double restriction digest (see 7.2.1.7), fractionated by agarose gel electrophoresis and Southern blotted as for plant genomic DNA (see 7.2.1.11). This extraction procedure yields sufficient DNA for 2 tracks on a Southern blot.

7.2.8 *Arabidopsis* Transformation and Tissue Culture Techniques

7.2.8.1 Transformation of *Arabidopsis thaliana* Root Explants

Adapted from the method of Valvekens *et al.* (1988).

All plant media are described in section 7.1.5. Manipulations were carried out in a flow hood using sterilized equipment and solutions and standard sterile technique. Growth conditions in the tissue culture room were 20°C with a 16 hour photoperiod of light intensity 50 μ E m⁻²s⁻¹.

To produce root material for *Agrobacterium*-mediated transformation *Arabidopsis thaliana* var. Landsberg *erecta* seeds, either wildtype or *co-2tt4* mutant, were surface sterilized as described in section 7.2.8.3. For each DNA construct 5 x 250ml conical flasks were filled with 50ml liquid IM and sterilized by autoclaving. 10-15 sterilized seeds were placed in each flask and incubated on a platform with slow shaking in a growth room for 2-3 weeks. Two days prior to transformation a culture of *Agrobacteria* containing the desired construct was initiated by inoculating 10ml LB rif₂₀₀ tet₁ with 10 μ l from the appropriate glycerol stock. This was grown at 30°C until day 1 of transformation when it was spun at 2000rpm for 10 minutes and the pellet resuspended in LB without antibiotics. The OD₆₀₀ was measured and an aliquot diluted to OD₆₀₀ = 0.1 in sterile liquid IM (usually 100ml).

Day 1; The root material was separated from all green tissue, pooled, chopped into 0.5cm lengths and placed into a plastic sieve within a petri dish. Roots from one of the flasks were treated with IM alone (no bacteria) to serve as an untransformed control. 25-50ml of the IM/*Agrobacteria* solution was poured over the root explants and agitated for 2 minutes. The contents of the sieve were drained, blotted on several changes of sterile Whatman 3MM paper and the roots spread thinly over the surface of CIM plates. These were sealed

(Micropore tape, 3M Medical-Surgical Division) and left to co-cultivate in the tissue culture room for 2 days.

Day 3; The roots were scraped off the surface of the plates, placed in a plastic sieve and washed with several changes of liquid IM until no longer turbid with bacteria. The roots were drained and blotted as before and divided into small clumps. These were spread thinly on the surface of SIM plates so that the diameter of each explant was no larger than 1cm. The plates were sealed and incubated as before. The untransformed roots were divided between 2 plates, with and without kanamycin to test for regeneration and kanamycin selection.

The explants were transferred to fresh SIM plates weekly. After about 10 days small green calli appeared all over the untransformed control explants on plates without kanamycin. A few days later calli were produced on the experimental plates. These produced plantlets which were transferred to GM plates to grow. Any additional callus material on the plantlets was trimmed off weekly. When the developing plants bolted they were transferred into a Magenta GA7 pot to set seed, the seeds harvested and sown directly onto GM/Km₅₀ plates to record the segregation ratio of resistant:sensitive seedlings. The progeny of the Km resistant plants were sown on GM/Km₅₀ plates to identify individuals homozygous for the transgene.

7.2.8.2 Transformation of *Arabidopsis thaliana* by Vacuum Infiltration

Based on the technique described by Bechtold *et al.* (1993) with modifications by Andrew Bent (Berkeley).

i) Plant material; Plants were grown in 4-inch pots in "*Arabidopsis*" mix (7.2.9(i)) rounded on top to form a dome. The pots were covered with mesh which was taped tightly down. They were watered, allowed to drain and Landsberg *erecta* seeds sprinkled over the top. Eight 4-inch pots were prepared for each construct, placed into an outer seed tray in short

days (SDs) and covered with a propagator lid for the first week. When the seedlings had produced a few true leaves they were thinned out to 10 plants per pot. After 3 weeks in SDs they were transferred to the long day greenhouse to allow the plants to bolt. When the bolts were approximately 10cm tall they were cut off near the base of the stem to encourage the plant to produce axillary shoots. Vacuum infiltration took place when the new shoots had produced their first 1 or 2 siliques.

ii) Agrobacteria culture; A 10ml culture of the appropriate strain was prepared 3 days in advance of infiltration (see 7.2.8.1). After 2 days the 10ml culture was used to inoculate 400ml LB rif₂₀₀ tet₁ in a 1L flask and grown for 24 hours at 30°C. The OD₆₀₀ was measured and if it was >1 the cells were harvested by centrifugation at 5000rpm for 10 minutes in a Sorvall H6000 rotor. The pellet was resuspended in sterile VIM (see 7.1.5) at an OD₆₀₀ of 0.8 (usually 3 x 400ml bottles of VIM per 400ml culture).

iii) Vacuum infiltration; Before infiltration 0.044µM benzylaminopurine (Sigma) and 0.03% Silwet L-77 (Union Carbide) were added to the VIM/*Agrobacteria* mixture. 400ml was poured into a 2L beaker and a single pot placed upside-down in the medium so that all plant stems were submerged. The beaker was placed in a vacuum dessicator (attached via a Refrigerated Vapour Trap RVT 100 to a Savant vacuum pump) and a vacuum applied for 30 seconds-3 minutes. The vacuum was released, the pot removed, rinsed in water and placed on its side on multiple layers of blueroll to drain. The VIM mixture was reused and replenished after every 3rd pot. When all pots had been treated the plants were covered with plastic autoclave bags and allowed to recover overnight. The next day they were uncovered and placed upright in the long day greenhouse where they remained until the plants set seed. Once the siliques started to turn brown all the plants on each pot were bulk bagged in 220mmx220mm cellophane bags and the seeds allowed to dry out. The seeds were harvested in bulk from each pot, separated from chaff, weighed and sterilized. After sterilization in 15ml falcon tubes batches of approximately 5000 seeds were resuspended

in 8ml molten GM K50 at 42°C and poured over the surface of 150mm GM K50 R20 plates (containing 20µg/ml Rovral fungicide). Seeds from different pots were never combined, so that independent transformants could be more easily identified. The plates were placed at 4°C for 2 days to break dormancy and transferred to a growth room. Kanamycin resistant plants were visible after about 7 days. When they had produced a few leaves they were transferred into soil to set seed.

7.2.8.3 Seed Sterilization

Seeds were surface sterilized before sowing on all plant growth media. To sterilize multiple batches of seeds together, seeds (usually 50-100) were enclosed in envelopes constructed from 3cm² pieces of Miracloth (Cal-Biochem) secured using paper clips. The envelopes were agitated in 70% ethanol for 2 minutes, bleach solution (5% NaOCL, 0.5% SDS) for 15 minutes followed by at least 4 rinses in sterile water. The envelopes were opened, the seeds allowed to dry for 2 hours in a sterile flow hood before being scattered onto petri dishes containing the appropriate medium (usually GM with or without Km). The plates were sealed and placed in a fridge for 2-5 days to break dormancy before being moved into the growth room. If a large amount of just one line of seeds was required (for example, to set up multiple flasks to grow root material for transformation) or after vacuum infiltration when 3-5000 seeds were to be sown per 150mm plate seeds were poured into 15ml Falcon tubes and sterilized using the method described above.

7.2.8.4 Selective Assay on Kanamycin

Seeds for selection on kanamycin were surface sterilized, sown on GM K50 (50µg/ml kanamycin) and scored for segregation of resistant to sensitive seedlings after 10 days. Sensitive seedlings produce cotyledons which bleach completely as they expand and

develop no true leaves and only stunted roots. Resistant seedlings grow as wildtype without kanamycin, and can be rescued onto GM plates or soil.

7.2.9 Plant Growth Conditions

i) Soil: In the glasshouses and controlled environment rooms (CER) plants were grown on a loam-based soil consisting of John Innes Potting Compost Number 1, vermiculite and grit in the ratio 1.5:1:1 by volume, or a mixture of Levington's Potting Compost Number 3 and grit in the ratio 6:1 by volume. Seeds were sown onto the surface of prewatered soil. Plants to be analysed phenotypically were usually sown in 4 inch pots, covered with clingfilm to maintain high humidity and transferred to the appropriate glasshouse or CER until sturdy enough to be pricked out into individual sections of P40 seed trays. After pricking out they were again covered with clingfilm for one week. Thereafter the clingfilm was removed and the plants watered once or twice a day as required. Seeds for plants to be grown only for DNA or RNA extraction were sown directly into seed trays and covered with cling film for a week.

ii) Stratification: Seed dormancy was broken by stratification: incubation at 4°C in the dark on damp filter paper for 2-7 days prior to transfer to the glasshouse or CER.

iii) Light conditions: In the summer plants in the glasshouses were grown in natural daylight. In the winter supplementary light was provided so that the minimum day length was 16 hours. Flowering time was measured under defined conditions by growing plants in Sanyo Gallenkamp Controlled Environment Rooms at 20°C with a relative humidity of 70%. Short days (SDs) comprised a photoperiod of 10 hours lit with 400W metal Halide Power Star lamps supplemented with 100W Tungsten Halide lamps. This provided a level of photosynthetically-active radiation of $113.7 \mu\text{moles photons m}^{-2}\text{s}^{-1}$ and a red:far red ratio of 2.41. Long days (LDs) comprised the same lighting regime with a 6 hour extension of the Tungsten Halide lamps.

iv) Out-crossing: Using a dissecting microscope and fine forceps a flower from the plant designated as the "female" was stripped of sepals and petals, and emasculated by removal of the immature anthers. Once a few (1-4) flowers on an inflorescence had been emasculated, mature flowers with dehiscent anthers and those already forming siliques were removed to ensure that seeds from self-pollinated flowers did not contaminate the out-crossed seeds. The emasculated flowers were then pollinated by dusting pollen from the designated "male" plant onto the stigma of the female. Pollination was repeated the following day, and new "female" flowers emasculated for pollination every 2 days as they developed. Once sufficient out-crossed siliques were developing the remaining immature flowers on the inflorescence were removed. Seeds were collected by enclosing the stem in a gas-permeable cellophane bag before the siliques ripened and split open. Once they had been harvested they were allowed to dry out at room temperature for > 2 weeks before sowing.

Appendix

One-Way Analysis of Variance (ANOVA) of flowering time of transformants in Long days (Chapter 5)

Transformant

0 = Landsberg erecta
 1 = Tn101
 2 = Tn104
 3 = Tn112
 4 = Tn114
 5 = Tn204
 6 = Tn206
 7 = Tn211
 8 = Tn213
 9 = Tn215
 10 = Tn33a
 11 = Tn43
 12 = Tn75a
 13 = TnG209
 14 = TnG210

Analysis of Variance on days to flowering

Source	DF	SS	MS	F	p
variety/	14	1000.50	71.46	32.46	0.000 ***
Error	269	592.21	2.20		
Total	283	1592.71			

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
0	10	20.100	1.449
1	18	22.333	1.534
2	20	18.400	1.465
3	20	18.900	1.410
4	20	17.700	0.733
5	17	20.235	2.047
6	20	20.400	1.847
7	20	19.800	2.462
8	20	20.700	0.923
9	20	19.500	1.821
10	20	18.900	1.410
11	20	19.950	1.468
12	20	21.300	0.923
13	20	14.200	0.616
14	19	21.316	0.946

Pooled StDev = 1.484

15.0 18.0 21.0 24.0

Tukey's pairwise comparisons (0.1%)

Family error rate = 0.100

Individual error rate = 0.00176

Critical value = 4.47

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-4.083 -0.384						
2	-0.116 3.516	2.410 5.457					
3	-0.616 3.016	1.910 4.957	-1.983 0.983				
4	0.584 4.216	3.110 6.157	-0.783 2.183	-0.283 2.683			
5	-2.004 1.734	0.512 3.684	-3.382 -0.288	-2.882 0.212	-4.082 -0.988		
6	-2.116 1.516	0.410 3.457	-3.483 -0.517	-2.983 -0.017	-4.183 -1.217	-1.712 1.382	
7	-1.516 2.116	1.010 4.057	-2.883 0.083	-2.383 0.583	-3.583 -0.617	-1.112 1.982	-0.883 2.083
8	-2.416 1.216	0.110 3.157	-3.783 -0.817	-3.283 -0.317	-4.483 -1.517	-2.012 1.082	-1.783 1.183
9	-1.216 2.416	1.310 4.357	-2.583 0.383	-2.083 0.883	-3.283 -0.317	-0.812 2.282	-0.583 2.383

10	-0.616 3.016	1.910 4.957	-1.983 0.983	-1.483 1.483	-2.683 0.283	-0.212 2.882	0.017 2.983
11	-1.666 1.966	0.860 3.907	-3.033 -0.067	-2.533 0.433	-3.733 -0.767	-1.262 1.832	-1.033 1.933
12	-3.016 0.616	-0.490 2.557	-4.383 -1.417	-3.883 -0.917	-5.083 -2.117	-2.612 0.482	-2.383 0.583
13	4.084 7.716	6.610 9.657	2.717 5.683	3.217 6.183	2.017 4.983	4.488 7.582	4.717 7.683
14	-3.048 0.616	-0.525 2.560	-4.418 -1.413	-3.918 -0.913	-5.118 -2.113	-2.646 0.485	-2.418 0.587
	7	8	9	10	11	12	13
8	-2.383 0.583						
9	-1.183 1.783	-0.283 2.683					
10	-0.583 2.383	0.317 3.283	-0.883 2.083				
11	-1.633 1.333	-0.733 2.233	-1.933 1.033	-2.533 0.433			
12	-2.983 -0.017	-2.083 0.883	-3.283 -0.317	-3.883 -0.917	-2.833 0.133		
13	4.117 7.083	5.017 7.983	3.817 6.783	3.217 6.183	4.267 7.233	5.617 8.583	
14	-3.018 -0.013	-2.118 0.887	-3.318 -0.313	-3.918 -0.913	-2.868 0.137	-1.518 1.487	-8.618 -5.613

Tukey's pairwise comparisons(1%)

Family error rate = 0.0100

Individual error rate = 0.000146

Critical value = 5.45

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-4.489 0.022						
2	-0.515 3.915	2.076 5.791					
3	-1.015 3.415	1.576 5.291	-2.308 1.308				
4	0.185 4.615	2.776 6.491	-1.108 2.508	-0.608 3.008			
5	-2.414 2.143	0.164 4.032	-3.722 0.051	-3.222 0.551	-4.422 -0.649		
6	-2.515 1.915	0.076 3.791	-3.808 -0.192	-3.308 0.308	-4.508 -0.892	-2.051 1.722	
7	-1.915 2.515	0.676 4.391	-3.208 0.408	-2.708 0.908	-3.908 -0.292	-1.451 2.322	-1.208 2.408
8	-2.815 1.615	-0.224 3.491	-4.108 -0.492	-3.608 0.008	-4.808 -1.192	-2.351 1.422	-2.108 1.508
9	-1.615 2.815	0.976 4.691	-2.908 0.708	-2.408 1.208	-3.608 0.008	-1.151 2.622	-0.908 2.708
10	-1.015 3.415	1.576 5.291	-2.308 1.308	-1.808 1.808	-3.008 0.608	-0.551 3.222	-0.308 3.308
11	-2.065 2.365	0.526 4.241	-3.358 0.258	-2.858 0.758	-4.058 -0.442	-1.601 2.172	-1.358 2.258
12	-3.415 1.015	-0.824 2.891	-4.708 -1.092	-4.208 -0.592	-5.408 -1.792	-2.951 0.822	-2.708 0.908
13	3.685 8.115	6.276 9.991	2.392 6.008	2.892 6.508	1.692 5.308	4.149 7.922	4.392 8.008
14	-3.450 1.018	-0.863 2.898	-4.748 -1.084	-4.248 -0.584	-5.448 -1.784	-2.989 0.828	-2.748 0.916

	7	8	9	10	11	12	13
8	-2.708 0.908						
9	-1.508 2.108	-0.608 3.008					
10	-0.908 2.708	-0.008 3.608	-1.208 2.408				
11	-1.958 1.658	-1.058 2.558	-2.258 1.358	-2.858 0.758			
12	-3.308 0.308	-2.408 1.208	-3.608 0.008	-4.208 -0.592	-3.158 0.458		
13	3.792 7.408	4.692 8.308	3.492 7.108	2.892 6.508	3.942 7.558	5.292 8.908	
14	-3.348 0.316	-2.448 1.216	-3.648 0.016	-4.248 -0.584	-3.198 0.466	-1.848 1.816	-8.948 -5.284

Tukey's pairwise comparisons (5%)

Family error rate = 0.0500

Individual error rate = 0.000800

Critical value = 4.80

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-4.220 -0.247						
2	-0.250 3.650	2.297 5.570					
3	-0.750 3.150	1.797 5.070	-2.093 1.093				
4	0.450 4.350	2.997 6.270	-0.893 2.293	-0.393 2.793			
5	-2.142 1.872	0.395 3.801	-3.497 -0.174	-2.997 0.326	-4.197 -0.874		
6	-2.250 1.650	0.297 3.570	-3.593 -0.407	-3.093 0.093	-4.293 -1.107	-1.826 1.497	
7	-1.650 2.250	0.897 4.170	-2.993 0.193	-2.493 0.693	-3.693 -0.507	-1.226 2.097	-0.993 2.193
8	-2.550 1.350	-0.003 3.270	-3.893 -0.707	-3.393 -0.207	-4.593 -1.407	-2.126 1.197	-1.893 1.293
9	-1.350 2.550	1.197 4.470	-2.693 0.493	-2.193 0.993	-3.393 -0.207	-0.926 2.397	-0.693 2.493
10	-0.750 3.150	1.797 5.070	-2.093 1.093	-1.593 1.593	-2.793 0.393	-0.326 2.997	-0.093 3.093
11	-1.800 2.100	0.747 4.020	-3.143 0.043	-2.643 0.543	-3.843 -0.657	-1.376 1.947	-1.143 2.043
12	-3.150 0.750	-0.603 2.670	-4.493 -1.307	-3.993 -0.807	-5.193 -2.007	-2.726 0.597	-2.493 0.693
13	3.950 7.850	6.497 9.770	2.607 5.793	3.107 6.293	1.907 5.093	4.374 7.697	4.607 7.793
14	-3.183 0.752	-0.639 2.674	-4.529 -1.302	-4.029 -0.802	-5.229 -2.002	-2.762 0.601	-2.529 0.698
	7	8	9	10	11	12	13
8	-2.493 0.693						
9	-1.293 1.893	-0.393 2.793					
10	-0.693 2.493	0.207 3.393	-0.993 2.193				
11	-1.743 1.443	-0.843 2.343	-2.043 1.143	-2.643 0.543			

12	-3.093 0.093	-2.193 0.993	-3.393 -0.207	-3.993 -0.807	-2.943 0.243		
13	4.007 7.193	4.907 8.093	3.707 6.893	3.107 6.293	4.157 7.343	5.507 8.693	
14	-3.129 0.098	-2.229 0.998	-3.429 -0.202	-4.029 -0.802	-2.979 0.248	-1.629 1.598	-8.729 -5.502

One-Way Analysis of Variance of leaf number of transformants in Long days (Chapter 5)

Transformant

0 = Landsberg erecta
 1 = Tn101
 2 = Tn104
 3 = Tn112
 4 = Tn114
 5 = Tn204
 6 = Tn206
 7 = Tn211
 8 = Tn213
 9 = Tn215
 10 = Tn33a
 11 = Tn43
 12 = Tn75a
 13 = TnG209
 14 = TnG210

Analysis of Variance on leaf number

Source	DF	SS	MS	F	p
variety/	14	193.619	13.830	16.40	0.000 ***
Error	269	226.786	0.843		
Total	283	420.405			

Individual 95% CIs For Mean Based on Pooled StDev

Level	N	Mean	StDev
0	10	7.200	0.919
1	18	9.778	1.003
2	20	7.550	0.759
3	20	7.300	0.801
4	20	8.050	0.826
5	17	8.471	1.007
6	20	8.050	1.146
7	20	7.550	0.999
8	20	6.950	0.999
9	20	6.950	0.826
10	20	7.000	0.649
11	20	6.400	0.995
12	20	7.850	0.988
13	20	6.400	0.598
14	19	7.895	1.100

Pooled StDev = 0.918

6.0 7.2 8.4 9.6

Tukey's pairwise comparisons (0.1%)

Family error rate = 0.100

Individual error rate = 0.00176

Critical value = 4.47

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-3.7224 -1.4331						
2	-1.4740 0.7740	1.2849 3.1707					
3	-1.2240 1.0240	1.5349 3.4207	-0.6677 1.1677				
4	-1.9740 0.2740	0.7849 2.6707	-1.4177 0.4177	-1.6677 0.1677			
5	-2.4272 -0.1140	0.3257 2.2887	-1.8780 0.0368	-2.1280 -0.2132	-1.3780 0.5368		
6	-1.9740 0.2740	0.7849 2.6707	-1.4177 0.4177	-1.6677 0.1677	-0.9177 0.9177	-0.5368 1.3780	
7	-1.4740 0.7740	1.2849 3.1707	-0.9177 0.9177	-1.1677 0.6677	-0.4177 1.4177	-0.0368 1.8780	-0.4177 1.4177
8	-0.8740 1.3740	1.8849 3.7707	-0.3177 1.5177	-0.5677 1.2677	0.1823 2.0177	0.5632 2.4780	0.1823 2.0177
9	-0.8740 1.3740	1.8849 3.7707	-0.3177 1.5177	-0.5677 1.2677	0.1823 2.0177	0.5632 2.4780	0.1823 2.0177
10	-0.9240 1.3240	1.8349 3.7207	-0.3677 1.4677	-0.6177 1.2177	0.1323 1.9677	0.5132 2.4280	0.1323 1.9677
11	-0.3240 1.9240	2.4349 4.3207	0.2323 2.0677	-0.0177 1.8177	0.7323 2.5677	1.1132 3.0280	0.7323 2.5677

12	-1.7740 0.4740	0.9849 2.8707	-1.2177 0.6177	-1.4677 0.3677	-0.7177 1.1177	-0.3368 1.5780	-0.7177 1.1177
13	-0.3240 1.9240	2.4349 4.3207	0.2323 2.0677	-0.0177 1.8177	0.7323 2.5677	1.1132 3.0280	0.7323 2.5677
14	-1.8286 0.4391	0.9285 2.8376	-1.2745 0.5850	-1.5245 0.3350	-0.7745 1.0850	-0.3930 1.5447	-0.7745 1.0850
	7	8	9	10	11	12	13
8	-0.3177 1.5177						
9	-0.3177 1.5177	-0.9177 0.9177					
10	-0.3677 1.4677	-0.9677 0.8677	-0.9677 0.8677				
11	0.2323 2.0677	-0.3677 1.4677	-0.3677 1.4677	-0.3177 1.5177			
12	-1.2177 0.6177	-1.8177 0.0177	-1.8177 0.0177	-1.7677 0.0677	-2.3677 -0.5323		
13	0.2323 2.0677	-0.3677 1.4677	-0.3677 1.4677	-0.3177 1.5177	-0.9177 0.9177	0.5323 2.3677	
14	-1.2745 0.5850	-1.8745 -0.0150	-1.8745 -0.0150	-1.8245 0.0350	-2.4245 -0.5650	-0.9745 0.8850	-2.4245 -0.5650

Tukey's pairwise comparisons (1%)

Family error rate = 0.0100

Individual error rate = 0.000146

Critical value = 5.45

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-3.9734 -1.1822						
2	-1.7204 1.0204	1.0782 3.3774					
3	-1.4704 1.2704	1.3282 3.6274	-0.8690 1.3690				
4	-2.2204 0.5204	0.5782 2.8774	-1.6190 0.6190	-1.8690 0.3690			
5	-2.6808 0.1396	0.1105 2.5039	-2.0879 0.2467	-2.3379 -0.0033	-1.5879 0.7467		
6	-2.2204 0.5204	0.5782 2.8774	-1.6190 0.6190	-1.8690 0.3690	-1.1190 1.1190	-0.7467 1.5879	
7	-1.7204 1.0204	1.0782 3.3774	-1.1190 1.1190	-1.3690 0.8690	-0.6190 1.6190	-0.2467 2.0879	-0.6190 1.6190
8	-1.1204 1.6204	1.6782 3.9774	-0.5190 1.7190	-0.7690 1.4690	-0.0190 2.2190	0.3533 2.6879	-0.0190 2.2190
9	-1.1204 1.6204	1.6782 3.9774	-0.5190 1.7190	-0.7690 1.4690	-0.0190 2.2190	0.3533 2.6879	-0.0190 2.2190
10	-1.1704 1.5704	1.6282 3.9274	-0.5690 1.6690	-0.8190 1.4190	-0.0690 2.1690	0.3033 2.6379	-0.0690 2.1690
11	-0.5704 2.1704	2.2282 4.5274	0.0310 2.2690	-0.2190 2.0190	0.5310 2.7690	0.9033 3.2379	0.5310 2.7690
12	-2.0204 0.7204	0.7782 3.0774	-1.4190 0.8190	-1.6690 0.5690	-0.9190 1.3190	-0.5467 1.7879	-0.9190 1.3190
13	-0.5704 2.1704	2.2282 4.5274	0.0310 2.2690	-0.2190 2.0190	0.5310 2.7690	0.9033 3.2379	0.5310 2.7690
14	-2.0771 0.6877	0.7192 3.0469	-1.4783 0.7888	-1.7283 0.5388	-0.9783 1.2888	-0.6055 1.7572	-0.9783 1.2888

	7	8	9	10	11	12	13
8	-0.5190 1.7190						
9	-0.5190 1.7190	-1.1190 1.1190					
10	-0.5690 1.6690	-1.1690 1.0690	-1.1690 1.0690				
11	0.0310 2.2690	-0.5690 1.6690	-0.5690 1.6690	-0.5190 1.7190			
12	-1.4190 0.8190	-2.0190 0.2190	-2.0190 0.2190	-1.9690 0.2690	-2.5690 -0.3310		
13	0.0310 2.2690	-0.5690 1.6690	-0.5690 1.6690	-0.5190 1.7190	-1.1190 1.1190	0.3310 2.5690	
14	-1.4783 0.7888	-2.0783 0.1888	-2.0783 0.1888	-2.0283 0.2388	-2.6283 -0.3612	-1.1783 1.0888	-2.6283 -0.3612

Tukey's pairwise comparisons (5%)

Family error rate = 0.0500

Individual error rate = 0.000800

Critical value = 4.80

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4	5	6
1	-3.8069 -1.3486						
2	-1.5570 0.8570	1.2153 3.2403					
3	-1.3070 1.1070	1.4653 3.4903	-0.7355 1.2355				
4	-2.0570 0.3570	0.7153 2.7403	-1.4855 0.4855	-1.7355 0.2355			
5	-2.5126 -0.0286	0.2532 2.3612	-1.9487 0.1075	-2.1987 -0.1425	-1.4487 0.6075		
6	-2.0570 0.3570	0.7153 2.7403	-1.4855 0.4855	-1.7355 0.2355	-0.9855 0.9855	-0.6075 1.4487	
7	-1.5570 0.8570	1.2153 3.2403	-0.9855 0.9855	-1.2355 0.7355	-0.4855 1.4855	-0.1075 1.9487	-0.4855 1.4855
8	-0.9570 1.4570	1.8153 3.8403	-0.3855 1.5855	-0.6355 1.3355	0.1145 2.0855	0.4925 2.5487	0.1145 2.0855
9	-0.9570 1.4570	1.8153 3.8403	-0.3855 1.5855	-0.6355 1.3355	0.1145 2.0855	0.4925 2.5487	0.1145 2.0855
10	-1.0070 1.4070	1.7653 3.7903	-0.4355 1.5355	-0.6855 1.2855	0.0645 2.0355	0.4425 2.4987	0.0645 2.0355
11	-0.4070 2.0070	2.3653 4.3903	0.1645 2.1355	-0.0855 1.8855	0.6645 2.6355	1.0425 3.0987	0.6645 2.6355
12	-1.8570 0.5570	0.9153 2.9403	-1.2855 0.6855	-1.5355 0.4355	-0.7855 1.1855	-0.4075 1.6487	-0.7855 1.1855
13	-0.4070 2.0070	2.3653 4.3903	0.1645 2.1355	-0.0855 1.8855	0.6645 2.6355	1.0425 3.0987	0.6645 2.6355
14	-1.9123 0.5228	0.8580 2.9081	-1.3431 0.6536	-1.5931 0.4036	-0.8431 1.1536	-0.4646 1.6163	-0.8431 1.1536
	7	8	9	10	11	12	13
8	-0.3855 1.5855						
9	-0.3855 1.5855	-0.9855 0.9855					
10	-0.4355 1.5355	-1.0355 0.9355	-1.0355 0.9355				

11	0.1645 2.1355	-0.4355 1.5355	-0.4355 1.5355	-0.3855 1.5855			
12	-1.2855 0.6855	-1.8855 0.0855	-1.8855 0.0855	-1.8355 0.1355	-2.4355 -0.4645		
13	0.1645 2.1355	-0.4355 1.5355	-0.4355 1.5355	-0.3855 1.5855	-0.9855 0.9855	0.4645 2.4355	
14	-1.3431 0.6536	-1.9431 0.0536	-1.9431 0.0536	-1.8931 0.1036	-2.4931 -0.4964	-1.0431 0.9536	-2.4931 -0.4964

One-Way Analysis of Variance of leaf number of pooled results from each construct in Long days (Chapter 5)

Construct
0 = Landsberg erecta
1 = 35S:co-3
2 = 35S:ZF
3 = 35S:co-7

Analysis of Variance on leaf number

Source	DF	SS	MS	F	p
constr	3	47.28	15.76	11.83	0.000 ***
Error	280	373.12	1.33		
Total	283	420.40			

				Individual 95% CIs For Mean Based on Pooled StDev			
Level	N	Mean	StDev	-----+-----+-----+-----			
0	10	7.200	0.919	(-----*-----)			
1	78	8.128	1.262			(---*---)	
2	97	7.567	1.145		(---*---)		
3	99	7.101	1.093	(---*---)			
Pooled StDev = 1.154				-----+-----+-----+-----			
				6.60	7.20	7.80	8.40

Tukey's pairwise comparisons (0.1%)
Family error rate = 0.100
Individual error rate = 0.0227
Critical value = 3.24
Intervals for (column level mean) - (row level mean)

	0	1	2
1	-1.817 -0.040		
2	-1.245 0.511	0.159 0.963	
3	-0.779 0.977	0.627 1.428	0.088 0.844

Tukey's pairwise comparisons (1%)
Family error rate = 0.0100
Individual error rate = 0.00204
Critical value = 4.40
Intervals for (column level mean) - (row level mean)

	0	1	2
1	-2.135 0.278		
2	-1.560 0.826	0.015 1.107	
3	-1.093 1.291	0.483 1.571	-0.047 0.979

Tukey's pairwise comparisons (5%)
Family error rate = 0.0500
Individual error rate = 0.0107
Critical value = 3.63
Intervals for (column level mean) - (row level mean)

	0	1	2
1	-1.923 0.067		
2	-1.351 0.617	0.111 1.012	
3	-0.884 1.082	0.579 1.476	0.043 0.889

One-Way Analysis of Variance of flowering time of pooled results from each construct in Long days (Chapter 5)

Construct

0 = Landsberg erecta
1 = 35S:co-3
2 = 35S:ZF
3 = 35S:co-7

Analysis of Variance on days to flowering

Source	DF	SS	MS	F	p
constr	3	60.65	20.22	3.69	0.012 *
Error	280	1532.07	5.47		
Total	283	1592.71			

Level	N	Mean	StDev	Individual 95% CIs For Mean Based on Pooled StDev
0	10	20.100	1.449	(-----+-----+-----+-----)
1	78	19.256	2.177	(-----*-----)
2	97	20.124	1.894	(-----*-----)
3	99	19.111	2.864	(-----*-----)

Pooled StDev = 2.339

19.20 20.00 20.80

Tukey's pairwise comparisons (0.1%)

Family error rate = 0.100

Individual error rate = 0.0227

Critical value = 3.24

Intervals for (column level mean) - (row level mean)

	0	1	2
1	-0.956 2.644		
2	-1.804 1.756	-1.682 -0.052	
3	-0.789 2.767	-0.666 0.957	0.247 1.778

Tukey's pairwise comparisons (1%)

Family error rate = 0.0100

Individual error rate = 0.00204

Critical value = 4.40

Intervals for (column level mean) - (row level mean)

	0	1	2
1	-1.601 3.288		
2	-2.441 2.393	-1.974 0.240	
3	-1.426 3.404	-0.957 1.247	-0.027 2.052

Tukey's pairwise comparisons (5%)

Family error rate = 0.0500

Individual error rate = 0.0107

Critical value = 3.63

Intervals for (column level mean) - (row level mean)

	0	1	2
1	-1.173 2.860		
2	-2.018 1.970	-1.780 0.046	
3	-1.003 2.981	-0.764 1.054	0.155 1.870

187

Tukey's pairwise comparisons (5%)

Family error rate = 0.0500

Individual error rate = 0.00452

Critical value = 4.12

Intervals for (column level mean) - (row level mean)

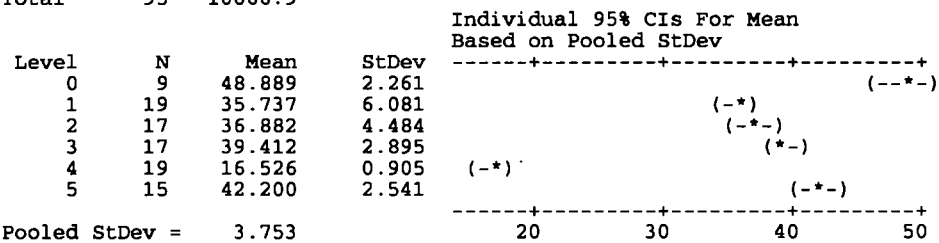
	0	1	2	3	4
1	3.640 13.132				
2	1.969 11.638	-5.497 2.333			
3	-1.972 7.697	-9.438 -1.608	-7.964 0.081		
4	21.272 30.763	13.827 21.437	15.298 23.129	19.240 27.070	
5	-1.545 8.345	-9.037 -0.935	-7.558 0.751	-3.617 4.692	-26.668 -18.567

One-Way Analysis of Variance of leaf number of 35S:co-7 transformants in short days (Chapter 5)

Transformant
0 = Landsberg erecta
1 = Tn33a
2 = Tn43
3 = Tn75a
4 = TnG209
5 = TnG210

Analysis of Variance on leaf number

Source	DF	SS	MS	F	p
variety/	5	9399.3	1879.9	133.47	0.000 ***
Error	90	1267.6	14.1		
Total	95	10666.9			



Tukey's pairwise comparisons (0.1%)
Family error rate = 0.100
Individual error rate = 0.00996
Critical value = 3.72
Intervals for (column level mean) - (row level mean)

	0	1	2	3	4
1	9.157 17.147				
2	7.937 16.076	-4.441 2.150			
3	5.408 13.547	-6.971 -0.379	-5.915 0.857		
4	28.368 36.357	16.008 22.413	17.060 23.652	19.590 26.181	
5	2.527 10.851	-9.873 -3.053	-8.815 -1.821	-6.285 0.709	-29.083 -22.264

Tukey's pairwise comparisons (1%)
Family error rate = 0.0100
Individual error rate = 0.000792
Critical value = 4.91
Intervals for (column level mean) - (row level mean)

	0	1	2	3	4
1	7.880 18.425				
2	6.635 17.378	-5.495 3.204			
3	4.106 14.848	-8.025 0.675	-6.999 1.940		
4	27.090 37.635	14.983 23.438	16.006 24.706	18.535 27.235	
5	1.195 12.183	10.964 -1.963	9.933 -0.702	-7.404 1.827	-30.174 -21.173

Tukey's pairwise comparisons (5%)

Family error rate = 0.0500

Individual error rate = 0.00452

Critical value = 4.12

Intervals for (column level mean) - (row level mean)

	0	1	2	3	4
1	8.728 17.576				
2	7.499 16.514	-4.796 2.505			
3	4.970 13.984	-7.325 -0.025	-6.279 1.221		
4	27.938 36.787	15.663 22.758	16.706 24.006	19.235 26.536	
5	2.079 11.299	-10.239 -2.687	-9.191 -1.445	-6.661 1.085	-29.450 -21.897

References

- Aasland R., Gibson T. J. and Stewart A. F. (1995) The PHD finger - implications for chromatin-mediated transcriptional regulation. *TIBS* **20** 56-59.
- Ahmad M. and Cashmore A. R. (1993) *HY4* gene of *A. thaliana* encodes a protein with characteristics of a blue light receptor. *Nature* **366** 162-166.
- Ahmad M. and Cashmore A. R. (1996) The *PEF* mutants of *Arabidopsis thaliana* define lesions early in the phytochrome signalling pathway. *Plant J.* **10** 1103-1110.
- Akbar A. (1987) Resynthesis of *Brassica napus* for earliness and daylength insensitivity. Ph.D. thesis, Swedish University of Agricultural Science, Sweden.
- Altschul S. F., Gish W., Miller W., Myers E. W. and Lipman D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215** 403-410.
- Alvarez J., Guli C. L., Yu X.-H. and Smyth D. R. (1992) *terminal flower*: a gene affecting inflorescence development in *Arabidopsis thaliana*. *Plant J.* **2** 103-116.
- Araki T., Kaya H., Kobayashi Y., Nakatani N. and Iwabuchi M. (1997) Flowering time gene *FT* encodes a *TFL* homolog. *8th International Conference on Arabidopsis Research, Madison, Wisconsin*, abstr. 5-1.
- Araki T. and Komeda Y. (1993a) Flowering in darkness in *Arabidopsis thaliana*. *Plant J.* **4** 801-811.
- Araki T. and Komeda Y. (1993b) Analysis of the role of the late-flowering locus, *GI*, in the flowering of *Arabidopsis thaliana*. *Plant J.* **3** 231-239.
- Arguello-Astorga G. R. and Herrera-Estrella L. R. (1996) Ancestral multipartite units in light-responsive plant promoters have structural features correlating with specific phototransduction pathways. *Plant Physiol.* **112** 1151-1166.
- Aronel V., Lemieux B., Hwang I., Gibson S., Goodman H. and Somerville C. R. (1992) Map-based cloning of a gene controlling Omega-3 fatty acid desaturation in *Arabidopsis*. *Science* **258** 1353-1355.
- Arumuganathan K. and Earle E. D. (1991) Nuclear DNA content of some important plant species. *Plant Mol. Biol. Rep.* **9** 208-218.
- Aukerman M. J. and Amasino R. M. (1996) Molecular genetic analysis of flowering time in *Arabidopsis*. *Sem. Cell Dev. Biol.* **7** 427-433.
- Bagnall D. (1993) Light quality and vernalization interact in controlling late-flowering in *Arabidopsis* ecotypes and mutants. *Ann. Bot.* **71** 75-83.
- Ballario P., Vittorioso P., Magrelli A., Talora C., Cabbibo A. and Macino G. (1996) White collar-1, a central regulator of blue-light responses in *Neurospora*, is a zinc-finger protein. *EMBO J.* **15** 1650-1657.
- Baltz R., Domon C., Pillay D. T. N. and Steinmetz A. (1992a) Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein. *Plant J.* **2** 713-721.
- Baltz R., Evrard J.-L., Bourdon V. and Steinmetz A. (1996) The pollen-specific LIM protein PLIM-1 from sunflower binds nucleic acids in vitro. *Sex. Plant Reprod.* **9** 264-268.
- Bechtold N., Ellis J. and Pelletier G. (1993) *In Planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. *C. R. Acad. Sci. Paris, Sciences de la vie/Life sciences* **316** 1194-1199.
- Bell C. J. and Ecker J. R. (1994) Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. *Genomics* **19** 137-144.

- Bellefroid E. J., Lecocq P. J., Benhida A. Poncelet D. A., Belayew A. and Martial J. A. (1989) The human genome contains hundreds of genes coding for finger proteins of the Krüppel type. *DNA* **8** 377-387.
- Bellmann R. and Werr W. (1992) Zmhox 1a, the product of a novel maize homeobox gene, interacts with the *Shrunken* 26 bp *feedback* control element. *EMBO J.* **11** 3367-3374.
- Bernier G. (1988) The control of floral evocation and morphogenesis. *Ann Rev. Physiol. Plant Mol. Biol.* **39** 179-219.
- Bernier G., Havelange A., Houssa C., Petitjean A. and Lejeune P. (1993) Physiological signals that induce flowering. *Plant Cell* **5** 1147-1155.
- Bernier G., Kinet J. M. and Sachs R. M. (1981) The Physiology of Flowering. Vol I (Boca Raton FL,: CRC Press).
- Bevan M. and The EU Arabidopsis Genome Project (1998) Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature* **391** 485-488.
- Beveridge C. A. and Murfet I. C. (1996) The *gigas* mutant in pea is deficient in the floral stimulus. *Physiol. Plant.* **96** 637-645.
- Birnboim H. C. and Doly J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* **7** 1513-1523.
- Blázquez M. A., Soowal L. N., Lee I. and Weigel D. (1997) *LEAFY* expression and flower initiation in *Arabidopsis*. *Development* **124** 3835-3844.
- Bleecker A. B., Estelle M. A., Somerville C. and Kende H. (1988) Insensitivity to ethylene conferred by a dominant mutation in *Arabidopsis thaliana*. *Science* **241** 1086-1089.
- Boylan M., Douglas N. and Quail P.H. (1994) Dominant negative suppression of *Arabidopsis* photoresponses by mutant phytochrome A sequences identifies spatially discrete regulatory domains in the photoreceptor. *The Plant Cell* **6** 449-460.
- Brendel V. and Karlin S. (1989) Association of charge clusters with functional domains of cellular transcription factors. *Proc. Natl. Acad. Sci. USA* **86** 5698-5702.
- Bruhn L., Hwang-Shum J.-J. and Sprague G. F., Jr. (1992) The N-terminal 96 residues of MCM1, a regulator of cell type-specific genes in *Saccharomyces cerevisiae*, are sufficient for DNA-binding, transcriptional activation and interaction with Alpha-1. *Mol. Cell. Biol.* **12** 3563-3572.
- Bullock W. O., Fernandez J. M. and Short J. M. (1987) XL1-blue - A high-efficiency plasmid transforming *recA* *Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5** 376.
- Burke D. T., Carle G. and Olson M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236** 806-812.
- Burn J. E., Bagnall D. J., Metzger J. D., Dennis E. S. and Peacock W. J. (1993a) DNA methylation, vernalization, and the initiation of flowering. *Proc. Natl. Acad. Sci. USA* **90** 287-291.
- Burn J. E., Smyth D. R., Peacock W. J. and Dennis E. S. (1993b) Genes conferring late flowering in *Arabidopsis thaliana*. *Genetica* **90** 147-155.
- Campos H., Giamoustaris A., Magrath R., Dean C. and Mithen R. (1996) Mapping of genes which regulate glucosinolate biosynthesis in *Brassica* and *Arabidopsis*, and progress towards map-based gene cloning. In: Plant Genome I, San Diego, California (abstract P175).
- Caspar T., Huber S. C. and Somerville C. R. (1985) Alterations in growth, photosynthesis and respiration in a starchless mutant of *Arabidopsis* (L.) deficient in chloroplast phosphoglucamutase activity. *Plant Physiol.* **79** 11-17.

- Caspar T., Lin T. -P., Kakefuda G., Benbow L., Preiss J. and Somerville C. R. (1991) Mutants of *Arabidopsis* with altered regulation of starch degradation. *Plant Physiol.* **95** 1181-1188.
- Castle L. A. and Sung Z. R. (1995) The *EMBRYONIC FLOWER* genes of *Arabidopsis* control shoot maturation. *Flowering Newsletter* **19** 12-19.
- Cavell A. C., Lydiate D. J., Parkin I. A. P., Dean C. and Trick M. (1998) Collinearity between a 30-centimorgan segment of *Arabidopsis thaliana* chromosome 4 and duplicated regions within the *Brassica napus* genome. *Genome* **41** 62-69.
- Chailakhayan M. Kh. (1936) New facts in support of the hormonal theory of plant development. *Dokl. Acad. Sci. USSR* **12** 443-447.
- Chandler J. and Dean C. (1994) Factors influencing the vernalization response and flowering time of late flowering mutants of *Arabidopsis thaliana* (L.) Heynh. *J. Exp. Bot.* **45** 1279-1288.
- Chandler J., Wilson A. and Dean C. (1996) *Arabidopsis* mutants showing an altered response to vernalization. *Plant J.* **10** 637-644.
- Chang C., Bowman J. L., DeJohn A. W., Lander E. S. and Meyerowitz E. M. (1988) Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **85** 6856-6860.
- Chang C., Kwok S. F., Bleeker A. B. and Meyerowitz E. M. (1993) *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. *Science* **262** 539-544.
- Chaudhury A. M., Letham S., Craig S. and Dennis E. S. (1993) *amp1* - a mutant with high cytokinin levels and altered embryonic pattern, faster vegetative growth, constitutive photomorphogenesis and precocious flowering. *Plant J.* **4** 907-916.
- Chen E. Y. and Seeburg P. H. (1985) Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* **4** 165-170.
- Chen L., Cheng J.-C., Castle L. and Sung R. (1997) *EMF* genes regulate *Arabidopsis* inflorescence development. *Plant Cell* **9** 2011-2024.
- Chory J. (1993) Out of the darkness: mutants reveal pathways controlling light-regulated development in plants. *TIG* **9** 167-172.
- Church G. M. and Gilbert W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA.* **81** 1991-1995.
- Clark J. H. and Dean C. (1994) Mapping *FRI*, a locus controlling flowering time and vernalisation response in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **242** 81-89.
- Clark S. E., Running M. P. and Meyerowitz E. M. (1993) *CLAVATA1*, a regulator of meristem and flower development in *Arabidopsis*. *Development* **119** 397-418.
- Clark S. E., Williams R. W. and Meyerowitz E. M. (1997) The *CLAVATA1* gene encodes a putative receptor kinase that controls shoot and floral meristem size in *Arabidopsis*. *Cell* **89** 575-585.
- Corbesier L., Gadiisseur I., Silvestre G., Jacquemard A. and Bernier G. (1996) Design in *Arabidopsis thaliana* of a synchronous system of floral induction by one long day. *Plant J.* **9** 947-952.
- Coupland G., Dash S., Goodrich J., Lee K., Long D., Puangsomlee P., Putterill J., Robson F., Sundberg E. and Wilson K. (1993) Molecular and genetic analysis of the control of flowering time in response to daylength in *Arabidopsis thaliana*. *Flowering Newsletter* **16** 27-32.
- Coupland G. (1995) Genetic and environmental control of flowering time in *Arabidopsis*. *TIG* **11** 393-397.

- Coupland G. (1997) Regulation of flowering by photoperiod in *Arabidopsis*. *Plant Cell Environ.* **20** 785-789.
- Courey A. J. and Tjian. R. (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. *Cell* **55** 887-898.
- Creusot F., Fouilloux E., Dron M., Lafleurriel J., Picard G., Billault A., Le Paslier D., Cohen D., Chaboute M-E., Durr A., Fleck J., Gigot C., Camilleri C., Bellini C., Caboche M. and Bouchez D. (1995) The CIC library: a large insert YAC library for genome mapping in *Arabidopsis thaliana*. *Plant J.* **8** 763-770.
- Crossley M, Merika M. and Orkin S. H. (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol. Cell. Biol.* **15** 2448-2456.
- Curtis O. F. and Chang H. T. (1930) The relative effectiveness of the temperature of the crown as contrasted with that of the rest of the plant upon the flowering of celery plants. *Am. J. Bot.* **17** 1047-1048.
- Damm B., Schmidt R. and Willmitzer L. (1989) Efficient transformation of *Arabidopsis thaliana* using direct gene transfer to protoplasts. *Mol. Gen. Genet.* **217** 6-12.
- Daniel-Vedele F. and Caboche M. (1993) A tobacco cDNA clone encoding a GATA-1 zinc-finger protein homologous to regulators of nitrogen-metabolism in fungi. *Mol. Gen. Genet.* **240** 365-373.
- Dawid I., Toyama R. and Taira M. (1995) LIM domain proteins. *C. R. Acad. Sci.* **318** 295-306.
- Dean C., Sjodin C., Page T., Jones J. D. G., and Lister C. (1992) Behaviour of the maize transposable element *Ac* in *Arabidopsis thaliana*. *The Plant J.* **2** 69-81.
- Deblaere R., Bytebier B., De Greve H., Deboeck F., Schell J., Van Montagu M. and Leemans J. (1985) Efficient octopine Ti plasmid-derived vectors for *Agrobacterium*-mediated gene transfer to plants. *Nuc. Acids Res.* **13** 4777-4788.
- Dehesh K., Hung H., Tepperman J. M. and Quail P. H. (1992) GT-2: a transcription factor with twin autonomous DNA-binding domains of closely related but different target sequence specificity. *EMBO J.* **11** 4131-4144.
- Dellaert L. M. W. (1980) The effects of dithiothreitol on radiation-induced genetic damage in *Arabidopsis thaliana* (L.) Heynh. *Mutation Res.* **71** 109-126.
- Deng X.-W., Matsui M., Wei N., Wagner D., Chu A. M., Feldman K. A. and Quail P. H. (1992) *COP1*, an *Arabidopsis* regulatory gene, encodes a protein with both a zinc-binding motif and G_β homologous domain. *Cell* **71** 791-801.
- Dennis E. S., Finnegan E. J., Bilodeau P., Chaudhury A., Genger R., Helliwell C. A., Sheldon C. C., Bagnall D. J. and Peacock W. J. (1996) Vernalization and the initiation of flowering. *Semin. Cell Dev. Biol.* **7** 441-448.
- De Paulis A., Sabatini S., De Pascalis L., Costantino P. and Capone I. (1996) A *rolB* regulatory factor belongs to a new class of single zinc finger plant proteins. *Plant J.* **10** 215-223.
- Dijkwel P. P., Huijser C., Weisbeek P. J., Chua N. and Smeeckens S. C. M. (1997) Sucrose control of phytochrome signalling in *Arabidopsis*. *Plant Cell* **9** 583-595.
- Ditta G., Stanfield S., Corbin D. and Helinski D. R. (1980) Broad host range DNA cloning systems for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* **77** 7347-7351.
- Ecker J. R. (1990) PFGE and YAC analysis of the *Arabidopsis* genome. *Methods* **1** 186-194.
- Ecker J. R. (1995) The ethylene signal transduction pathway in plants. *Science* **268** 667-675.

- Eimert K., Wang S.-M., Lue W.-L. and Chen J. (1995) Monogenic recessive mutations causing both late floral initiation and excess starch accumulation in *Arabidopsis*. *Plant Cell* **7** 1703-1712.
- Emmler K., Stockhaus J., Chua N.-H. and Schafer E. (1995) An amino-terminal deletion of rice phytochrome A results in a dominant negative suppression of tobacco phytochrome A activity in transgenic tobacco seedlings. *Planta* **197** 103-110.
- Eskins K. (1992) Light quality effects on *Arabidopsis* development. Red, blue and far red regulation of flowering and morphology. *Physiologia Plantarum* **86** 439-444.
- Estruch J. J., Crossland L. and Goff S. A. (1994) Plant activating sequences: positively charged peptides are functional as transcriptional activation domains. *Nuc. Acids Res.* **22** 3983-3989.
- Evans T. and Felsenfeld G. (1989) The erythroid-specific transcription factor Eryf1 - a new finger protein. *Cell* **58** 877-885.
- Feinbaum R. L. and Ausubel F. M. (1988) Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Mol. Cell. Biol.* **8** 1985-1992.
- Feinberg A. P. and Vogelstein P. M. (1983) A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. *Anal. Biochem.* **132** 6-13.
- Feldman K. A., Malmberg R. L. and Dean C. (1994) Mutagenesis in *Arabidopsis*. In: *Arabidopsis*. Meyerowitz E. M. and Somerville C. (eds.) (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) 137-172.
- Feldman K. A., Wierzbicki A. M., Reiter R. S. and Coomber S. A. (1991) T-DNA insertional mutagenesis in *Arabidopsis*. In: *Plant Molecular Biology*, Vol 2. Hermann R. and Larkins B. (eds.) (Plenum Press, NY) 563-574.
- Fortini M. E., Lai Z. and Rubin G. M. (1991) The *Drosophila zfh-1* and *zfh-2* genes encode novel proteins containing both zinc-finger and homeodomain motifs. *Mech. Dev.* **34** 113-122.
- Franken P., Schrell S., Peterson P. A., Saedler H. and Wienand U. (1994) Molecular analysis of protein domain function encoded by the *myb*-homologous maize genes *C1*, *Zm 1* and *Zm 38*. *Plant J.* **6** 21-30.
- Freemont P. S., Hanson I. M. and Trowsdale J. (1991) A novel cysteine-rich sequence motif. *Cell* **64** 483-484.
- Freydt G., Kim S. K. and Horvitz H. R. (1990) Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature* **344** 876-879.
- Frohman M. A., Dush M. K. and Martin G. R. (1988) Rapid production of full length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85** 8998-9002.
- Fu Y. H. and Marzluf G. H. (1990) *Nit-2*, the major nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell Biol.* **10** 1056-1065.
- Garcia-Bustos J., Heitman J. and Hall M. N. (1991) Nuclear protein localization. *Biochim. Biophys. Acta.* **1071** 83-101.
- Garner W. W. and Allard H. A. (1920) Effect of the relative length of day and night and other factors of the environment on growth and reproduction in plants. *J. Agric. Res.* **18** 553-606.
- Garner W. W. and Allard H. A. (1923) Further studies in photoperiodism, the response of the plant to relative length of day and night. *J. Agric. Res.* **23** 871-920.
- Giraudat J., Hauge B. M., Valon C., Smalle J., Parcy F. and Goodman H.M. (1992) Isolation of the *Arabidopsis ABI3* gene by positional cloning. *Plant Cell* **4** 1251-1261.

- Glover J. N. M. and Harrison S. C. (1995) Crystal structure of the heterodimeric bZIP transcription factor C-FOS-C-JUN bound to DNA. *Nature* **373** 257-261.
- Goff S. A., Cone K. C. and Fromm M. E. (1991) Identification of functional domains in the maize transcriptional activator C1: comparison of wild-type and dominant inhibitor proteins. *Genes and Development* **5** 298-309.
- Goldberg D. A. (1980) Isolation and partial characterization of the *Drosophila* alcohol dehydrogenase gene. *Proc. Natl. Acad. Sci. USA* **77** 5794-5798.
- Goodrich J., Puangsomlee P., Martin M., Long D., Meyerowitz E. M. and Coupland G. (1997) A polycomb group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* **386** 44-51.
- Goto N., Kumagai T. and Koornneef M. (1991) Flowering responses to light-breaks in photomorphogenic mutants of *Arabidopsis thaliana*, a long day plant. *Physiol. Plant.* **83** 209-215.
- Grbic V. and Bleecker A. B. (1996) An altered bodyplan is conferred on *Arabidopsis* plants carrying dominant alleles of two genes. *Development* **122** 2395-2403.
- Grbic V. and Gray J. (1997) *Aerial rosette 1*, *ART1*, is a new late flowering gene of *Arabidopsis thaliana*. *8th International Conference on Arabidopsis Research, Madison, Wisconsin*, abst. 5-9.
- Grill E. and Somerville C. (1991) Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Mol Gen Genet.* **226** 484-490.
- Guerineau F., Brooks L. and Mullineaux P. (1991) Effect of deletions in the cauliflower mosaic virus polyadenylation sequence on the choice of the polyadenylation sites in tobacco protoplasts. *Mol. Gen. Genet.* **226** 141-144.
- Guo H., Yang H., Mockler T. and Lin C. (1998) Regulation of flowering time by *Arabidopsis* photoreceptors. *Science* **279** 1360.
- Guzman P. and Ecker J. R. (1988) Development of large DNA methods for plants: molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. *Nucleic Acids Res.* **16** 11091-11105.
- Halliday K. J., Devlin P. F., Whitelam G. C., Hanhart C. J. and Koornneef M. (1996) The *ELONGATED* gene of *Arabidopsis* acts independently of light and gibberellins in the control of elongation growth. *Plant J.* **9** 305-312.
- Halliday K. J., Koornneef M. and Whitelam G. C. (1994) Phytochrome B and at least one other phytochrome mediates the accelerated flowering response of *Arabidopsis thaliana* L. to low red/far red ratio. *Plant Physiol.* **104** 1311-1315.
- Hamner K. C. and Bonner J. (1938) Photoperiodism in relation to hormones as factors in floral initiation and development. *Bot. Gaz.* **100** 388-431.
- Hanahan D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166** 557-580.
- Harrison S. C. (1991) A structural taxonomy of DNA-binding domains. *Nature* **353** 715-719.
- Hauge B. M. and Goodman H. M. (1992) Genome mapping in *Arabidopsis*. In: *Methods in Arabidopsis Research*. Koncz C., Chua N-H and Schell (eds.) World Scientific, Singapore, 191-224.
- Hauge B. M., Hanley S., Giraudat J. and Goodman H. (1991) Mapping the *Arabidopsis* genome. In: Jenkins G. I. and Schuch W. (eds.) *Molecular Biology of Plant Development*. (Company of Biologists, Cambridge) 45-56.
- Haughn G. W., Schultz E. A. and Martinez-Zapater J. M. (1995) The regulation of flowering in *Arabidopsis thaliana*: meristems, morphogenesis and mutants. *Can J. Bot.* **73** 959-981.

- Hawkins R. B. (1979) Double strand-breaks and DNA-to-protein cross-links induced by fast neutrons in bacteriophage DNA. *Int. J. Radiat. Biol.* **35** 1-13.
- Hedge I. C. (1976) A systematic and geographical survey of the old world Cruciferae. In: *The Biology and Chemistry of the Cruciferae*. Macleod A. J. and Jones B. M. J. (eds.) London :Academic Press, 1-46.
- Hemerly A., de Almeida Engler J., Bergounioux C., Van Montagu M., Engler G., Inzé D. and Ferreira P. (1995) Dominant negative mutants of the Cdc2 kinase uncouple cell division from iterative plant development. *The EMBO J.* **14** 3925-3936.
- Hempel F. D. and Feldman L. J. (1994) Bi-directional inflorescence development in *Arabidopsis thaliana*: acropetal initiation of flowers and basipetal initiation of paraclades. *Planta* **192** 276-286.
- Hempel F. D. and Feldman L. J. (1995) Specification of chimeric flowering shoots in wild-type *Arabidopsis*. *Plant J.* **8** 725-731.
- Herskowitz I. (1987) Functional inactivation of genes by dominant negative mutations. *Nature* **329** 219-222.
- Hicks K. A., Millar A. J., Carre I. A., Somers D. E., Straume M., Meeks-Wagner D. R. and Kay S. A. (1996a) Conditional circadian dysfunction of the *Arabidopsis early-flowering3* mutant. *Science* **274** 790-792.
- Hicks K. A., Sundas A. and Meeks-Wagner D. R. (1996b) *Arabidopsis* early flowering mutants reveal multiple levels of regulation in the vegetative-to-floral transition. *Sem. Cell Dev. Biol.* **7** 409-418.
- Hill A., Nantel A., Rock C. D. and Quatrano R. S. (1996) A conserved domain of the viviparous-1 gene product enhances the DNA-binding activity of the bZIP protein EMBP-1 and other transcription factors. *J. Biol. Chem.* **271** 3366-3374.
- Hofte H. *et al.* (1993) An inventory of 1,152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* **4** 1051-1061.
- Hollenberg M., Weinberger C., Ong E. S., Cerelli G., Lebo A. R., Thompson E. B., Rosenfeld M. G. and Evans R. M. (1985) Primary structure and expression of a functional human glucocorticoid receptor cDNA. *Nature* **318** 635-641.
- Holsters M., de Waele D., Depicker A., Messens E., van Montagu M. and Schell J. (1978) Transfection and transformation of *Agrobacterium tumefaciens*. *Molec. Gen. Genet.* **163** 181-187.
- Hope I. A. and Struhl K. (1986) Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* **46** 885-894.
- Hua J., Chang C., Sun Q. and Meyerowitz E. M. (1995) Ethylene insensitivity conferred by *Arabidopsis ERS* gene. *Science* **269** 1712-1714.
- Hussein H. A. S. (1968) Genetic analysis of mutagen-induced flowering time variation in *Arabidopsis thaliana* (L.) Heyhn. Ph.D. thesis. *Landbouwhogeschool Meded.* **68** 1-88.
- Ishiguro S. and Nakamura K. (1994) Characterization of a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and β -amylase from sweet potato. *Mol. Gen. Genet.* **244** 563-571.
- Jacobs G. H. (1992) Determination of base recognition positions of zinc fingers from sequence analysis. *EMBO J.* **11** 4507-4517.
- Jacobsen S. E. and Olszewski N. E. (1993) Mutations at the *SPINDLY* locus of *Arabidopsis* alter gibberellin signal transduction. *Plant Cell* **5** 887-896.
- Jansen R. C., Van Ooijen J. W., Stam P., Lister C. and Dean C. (1995) Genotype by environment interaction in genetic mapping of multiple quantitative trait loci. *Theor. Appl. Genet.* **91** 33-37.
- Johnson E., Bradley M., Harberd N. P. and Whitelam G. C. (1994) Photoresponses of light-grown PhyA mutants of *Arabidopsis*. Phytochrome A is required for the perception of daylength extensions. *Plant Physiol.* **105** 141-149.

- Jones J. D. G., Shlumukov L., Carland F., English J., Scofield S. R., Bishop G. J. and Harrison K (1992) Effective vectors for transformation, expression of heterologous genes, and assaying transposon excision in transgenic plants. *Transgenic Res.* **1** 285-297.
- Karlsson B. H., Sills G. R. and Nienhuis J. (1993) Effects of photoperiod and vernalisation on the number of leaves at flowering in 32 *Arabidopsis thaliana* (Brassicaceae) ecotypes. *Am J. Bot.* **80** 646-648.
- Karlsson O., Thor S., Norberg T., Ohlsson H. and Edlund T. (1990) Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* **344** 879-882.
- Kania T., Russenberger D., Peng S., Apel K. and Melzer S. (1997) *FPF1* promotes flowering in *Arabidopsis*. *Plant Cell* **9** 1327-1338.
- Keegan L., Gill G. and Ptashne M. (1986) Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231** 699-704.
- Keith D. J. (1995) Genetic analysis of quantitative traits in *Brassica napus*. Ph.D. thesis, University of East Anglia, UK.
- King R. and Bagnall D. (1996) Photoreceptors and the photoperiodic response controlling flowering of *Arabidopsis*. *Cell Dev. Biol.* **7** 449-454.
- Klimyuk V. I., Carroll B. J., Thomas C. M and Jones J. D. G. (1993) Alkali treatment for rapid preparation of plant material for reliable PCR analysis. *The Plant J.* **3** 493-494.
- Klug A. and Schwabe J. W. R. (1995) Zinc fingers. *FASEB J.* **9** 597-604.
- Knott J. E. (1934) Effect of a localized photoperiod on spinach. *Proc. Am. Soc. Hort. Sci* **31** 152-154.
- Koncz C., Kreuzaler F., Kálmán Zs. and Schell J (1984) A simple method to transfer, integrate and study expression of foreign genes, such as chicken ovalbumin and α -actin in plant tumours. *EMBO J.* **3** 1929-1937.
- Konieczny A. and Ausubel F. (1993) A procedure for quick mapping of *Arabidopsis* mutants using ecotype specific markers. *Plant J.* **4** 403-410.
- Koornneef M. (1987) Linkage map of *Arabidopsis thaliana* (2n = 10). In: O' Brien S. J. (ed.) Genetic Maps (Cold Spring Harbor Laboratories, Cold Spring Harbor, NY) 742-745.
- Koornneef M. (1990) Mutations affecting the testa color in *Arabidopsis*. *Arabid. Inf. Serv.* **27** 1-4.
- Koornneef M., Alonso-Blanco C., Blankestijn-de Vries H., Hanhart C. J. and Peeters A. J. M. (1998) Genetic interactions among late-flowering mutants of *Arabidopsis*. *Genetics* **148** 885-892.
- Koornneef M., Blankestijn-de Vries H., Hanhart C., Soppe W. and Peeters T. (1994) The phenotype of some late flowering mutants is enhanced by a locus on chromosome 5 that is not effective in the Landsberg *erecta* wild type. *Plant J.* **6** 911-919.
- Koornneef M., Dellaert L. W. M. and van der Veen J. H. (1982) EMS- and radiation-induced mutation frequencies at individual loci in *Arabidopsis thaliana* (L) Heynh. *Mut. Res.* **93** 109-123.
- Koornneef M., Hanhart C. J. and van der Veen J. H. (1991) A genetic and physiological analysis of late flowering mutants in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **229** 57-66.
- Koornneef M. and Peeters J. M. (1997) Floral transition mutants in *Arabidopsis*. *Plant Cell Environ.* **20** 779-784.

- Koornneef M., Rolff E. and Spruit C. J. P. (1980) Genetic control of light inhibition of hypocotyl elongation in *Arabidopsis thaliana* L. Heynh. *Zeitschrift Pflanzenphysiol.* **100** 147-160.
- Koornneef M. and van der Veen J. H. (1980) Induction and analysis of gibberellin-sensitive mutants in *Arabidopsis thaliana* (L.) Heynh. *Theor. Appl. Genet.* **58** 257-263.
- Koornneef M., van Eden J., Hanhart C. J., Stam P., Braaksma F. J. and Feenstra W. J. (1983) Linkage map of *Arabidopsis thaliana*. *J. Hered.* **74** 265-272.
- Korfhage U., Trezzini G. F., Meier I., Hahlbrock K. and Somssich I. E. (1994) Plant homeodomain protein involved in transcriptional regulation of a pathogen defense-related gene. *Plant Cell* **6** 695-708.
- Kowalski S. P., Lan T.-H., Feldmann K. A. and Paterson A. H. (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. *Genetics* **138** 499-510.
- Kudla B., Caddick M., Langdon T., Martinezrossi N. M., Bennett C. F., Sibley S., Davies R. W. and Arst H. N. (1990) The regulatory gene *areA* mediating nitrogen metabolite repression in *Aspergillus nidulans* - mutations affecting specificity of gene activation alter a loop residue of a putative zinc finger. *EMBO J.* **9** 1355-1364.
- Kumar V., Green S., Stack G., Berry M., Jin J.-R. and Chambon P. (1987) Functional domains of the human estrogen receptor. *Cell* **51** 941-951.
- Lagercrantz U. and Lydiate D. J. (1995) RFLP mapping in *Brassica nigra* indicates differing recombination rates in male and female meioses. *Genome* **38** 255-264.
- Lagercrantz U. and Lydiate D. J. (1996) Comparative genome mapping in *Brassica*. *Genetics* **144** 1903-1910.
- Lagercrantz U., Putterill J., Coupland G. and Lydiate D. (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. *Plant J.* **9** 13-20.
- Laibach F. (1907) Zur Frage nach der Individualität der Chromosomen im Pflanzenreich. *Beih. Biologisches Zentralblatt I. Abt.* **22** 191-210.
- Laibach F. (1943) *Arabidopsis thaliana* (L.) Heynh. als Objekt für genetische und entwicklungsphysiologische Untersuchungen. *Botanisches Archiv.* **44** 439-455.
- Laibach F. (1951) Über sommer- und winterannuelle Rassen von *Arabidopsis thaliana* (L.) Heynh. Ein Beitrag zur Ätiologie der Blütenbildung. *Beitr. Biol. Pflanzen* **28** 173-210.
- Lam E., Kano-Murakami Y., Gilmartin P., Niner B. and Chua N.-H. (1990) A metal-dependent DNA-binding protein interacts with a constitutive element of a light-responsive promoter. *Plant Cell* **2** 857-866.
- Lang A. (1952) Physiology of flowering. *Ann Rev Plant Physiol.* **3** 265-306.
- Ledger S. E., Dare A. P. and Putterill J. J. (1996) *COL2* (Accession No. L81119 and L81120) is a homologue of the *Arabidopsis* flowering-time gene *CONSTANS* (PGR96-081). *Plant Physiol.* **112** 862.
- Lee I., Bleecker A. and Amasino R. (1993) Analysis of naturally occurring late flowering in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **237** 171-176.
- Lee I., Michaels S. D., Masshardt A. S. and Amasino R. (1994a) The late flowering phenotype of *FRIGIDA* and mutations in *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. *Plant J.* **6** 903-909.
- Lee I., Aukerman M. J., Gore S. L., Lohman K. N., Michaels S. D., Weaver L. M., John M. C., Feldmann K. A. and Amasino R. M. (1994b) Isolation of *LUMINIDEPENDENS*: a gene involved in the control of flowering time in *Arabidopsis*. *Plant Cell* **6** 75-83.

- Leutweiler L. S., Hough-Evans B. R. and Meyerowitz E. M. (1984) The DNA of *Arabidopsis*. *Mol. Gen. Genet.* **194** 15-23.
- Levy Y. Y. and Dean C. (1998) Control of flowering time. *Curr. Op. Plant Biol.* **1** 49-54.
- Lightner J. and Caspar T. (1998) Seed mutagenesis of *Arabidopsis*. In: Methods in Molecular Biology: *Arabidopsis* Protocols. Martínez-Zapater J. M. and Salinas J. (eds.) (Humana Press, Totowa, New Jersey) 91-104.
- Lin C., Ahmad M., Chan A. R. and Cashmore A. R. (1996) CRY2: A second member of the *Arabidopsis* cryptochrome gene family. *Plant Physiol.* **110** 1047.
- Lin T. P., Caspar T., Somerville C. and Preiss J. (1988) Isolation and characterization of a starchless mutant of *Arabidopsis thaliana* (L.) Heynh lacking ADP glucose pyrophosphorylase. *Plant Physiol.* **86** 1131-1135.
- Linden H. and Macino G. (1997) White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*. *EMBO J.* **16** 98-109.
- Lippuner V., Cyert M. S. and Gasser C. S. (1996) Two classes of plant cDNA clones differentially complement yeast calcineurin mutants and increase salt tolerance of wild-type yeast. *J. Biol. Chem.* **271** 12859-12866.
- Lovering R., Hanson I. M., Borden K. L. B., Martin S., O'Reilly N. J., Evan G. I., Rahman D., Pappin D. J. C., Trowsdale J. and Freemont P. S. (1993) Identification and preliminary characterization of a protein motif related to the zinc finger. *Proc. Natl. Acad. Sci. USA* **90** 2112-2116.
- Lydiate D., Sharpe A., Lagercrantz U. and Parkin I. (1993) Mapping the *Brassica* genome. *Outlook Agric.* **22** 85-92.
- Ma P. C. M., Rould M., Weintraub H. and Pabo C. O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **77** 451-459.
- Mackay J. P. and Crossley M. (1998) Zinc fingers are sticking together. *TIBS* **23** 1-42.
- Macknight R., Bancroft I., Page T., Lister C., Schmidt R., Love K., Westphal L., Murphy G., Sherson S., Cobbett C. and Dean C. (1997) FCA, a gene controlling flowering time in *Arabidopsis*, encodes a protein containing RNA-binding domains. *Cell* **89** 737-745.
- Madueño F., Ruiz-Garcia L., Salinas J. and Martínez-Zapater J. M. (1996) Genetic interactions that promote the floral transition in *Arabidopsis*. *Sem. Cell Dev. Biol.* **7** 401-407.
- Mandel M. A. and Yanofsky M. F. (1995) A gene triggering flower formation in *Arabidopsis*. *Nature* **377** 522-524.
- Maniatis T., Fritsch E. F. and Sambrook J. (1989) Molecular Cloning: A Laboratory Manual. Second edition (Cold Spring Harbour Laboratory Press: New York).
- Martínez-García M., Garcidueñas-Piña C. and Guzmán P. (1996) Gene isolation in *Arabidopsis thaliana* by conditional overexpression of cDNAs toxic to *Saccharomyces cerevisiae*: identification of a novel early response zinc-finger gene. *Mol. Gen. Genet.* **252** 587-596.
- Martínez-Zapater J. M., Coupland G., Dean C. and Koornneef M. (1994) The transition to flowering in *Arabidopsis*. In Meyerowitz E. M. and Somerville C. R. (eds) *Arabidopsis* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY) 403-433.
- Martínez-Zapater J. M., Jarillo J. A., Cruz-Alvarez M., Roldán M. and Salinas J. (1995) *Arabidopsis* late-flowering *fve* mutants are affected in both vegetative and reproductive development. *Plant J.* **7** 543-551.
- McCubbin A. G., Chung Y.-Y. and Kao T.-h (1997) A mutant S₃ RNase of *Petunia inflata* lacking RNase activity has an allele-specific dominant negative effect on self-incompatibility interactions. *The Plant Cell* **9** 85-95.

- McKelvie A. D. (1962) A list of mutant genes in *Arabidopsis thaliana* (L.) Heynh. *Radiat. Bot.* **1** 233-241.
- McNellis T. W., Torii K. U. and Deng X.-W. (1996) Expression of an N-terminal fragment of COP1 confers a dominant-negative effect on light-regulated seedling development in *Arabidopsis*. *The Plant Cell* **8** 1491-1503.
- McNellis T. W., von Arnim A. G., Araki T., Komeda Y., Miséra S. and Deng X.-W. (1994) Genetic and molecular analysis of an allelic series of *cop1* mutants suggests functional roles for the multiple protein domains. *Plant Cell* **6** 487-500.
- Meisel L. and Lam E. (1996) The conserved ELK-homeodomain of KNOTTED-1 contains two regions that signal nuclear localization. *Plant Mol. Biol.* **30** 1-14.
- Meissner R. and Michael A. J. (1997) Isolation and characterization of a diverse family of *Arabidopsis* two and three-fingered C₂H₂ zinc finger protein genes and cDNAs. *Plant Mol. Biol.* **33** 615-624.
- Merika M. and Orkin S. H. (1995) Functional synergy and physical interactions of the erythroid transcription factor GATA-1 and Krüppel family proteins, Sp1 and EKLF. *Mol. Cell. Biol.* **15** 2437-2447.
- Meshi T. and Iwabuchi M. (1995) Plant transcription factors. *Plant Cell Physiol.* **36** 1405-1420.
- Messing J. (1979) A multipurpose cloning system based on single-stranded DNA bacteriophage M13. *Recomb. DNA Tech. Bull.* **2** 93-111.
- Michael A. J., Hofer J. M. I. and Ellis T. H. N. (1996) Isolation by PCR of a cDNA clone from pea petals with similarity to petunia and wheat zinc finger proteins. *Plant Mol. Biol.* **30** 1051-1058.
- Millar A. J. and Kay S. A. (1991) Circadian control of *cab* gene transcription and mRNA accumulation in *Arabidopsis*. *Plant Cell* **3** 541-550.
- Millar A. J., Straume M., Chory J., Chua N.-H. and Kay S. A. (1995) The regulation of circadian period by phototransduction pathways in *Arabidopsis*. *Science* **267** 1163-1166.
- Miller J. H. (1978) The *lacI* gene: its role in *lac* operon control and its use as a genetic system. In: *The Operon* (Eds. Miller J. H. and Reznikoff W. S. Cold Spring Harbour Press: New York) 31-38.
- Miller J., McLachlan A. D., and Klug A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J.* **4** 1609-1614.
- Mishima K., Tsuchiya M., Nightingale M. S., Moss J. and Vaughan M. (1993) ARD1, a 64-kDa guanine nucleotide binding protein with a carboxyl-terminal ADP-ribosylation factor domain. *J. Biol. Chem.* **268** 8801-8807.
- Mitchell P. J. and Tjian R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* **245** 371-378.
- Mizukami Y., Huang H., Tudor M., Hu Y. and Ma H. (1996) Functional domains of the floral regulator AGAMOUS: characterization of the DNA binding domain and analysis of dominant negative mutations. *The Plant Cell* **8** 831-845.
- Mizukami Y. and Ma H. (1997) Determination of *Arabidopsis* floral meristem identity by AGAMOUS. *Plant Cell* **9** 393-408.
- Mozley D. and Thomas B. (1995) Developmental and photobiological factors affecting photoperiodic induction in *Arabidopsis thaliana* Heynh. Landsberg *erecta*. *J Exp. Biol.* **46** 173-179.
- Muller H. J. (1932) Further studies on the nature and causes of gene mutations. In: *Proceedings of the Sixth International Congress of Genetics* (Ed. Jones D. F. Brooklyn Botanic Gardens, Menasha, Wisconsin) **1** 213-255.

- Murfet I. C. (1977) Environmental interaction and the genetics of flowering. *Ann. Rev. Plant Physiol.* **28** 253-278.
- Murphy D. J. (1992) Modifying oilseed crops for non-edible products. *Tibtech* **10** 84-87.
- Murphy L. and Scarth R. (1994) Vernalization response in spring oilseed rape (*Brassica napus* L.) cultivars. *Can J. Plant Sci.* **74** 275-277.
- Nam H.-G., Giraudat J., den Boer B., Moonan F., Loos W. D. B., Hauge B. M. and Goodman H. M. (1989) Restriction fragment length polymorphism linkage map of *Arabidopsis thaliana*. *Plant Cell* **1** 699-705.
- Napp-Zinn K. (1969) *Arabidopsis thaliana* (L) Heyhn. In: The Induction of Flowering. Some case histories (Evans L. T. eds.) Melbourne, Macmillan 291-304.
- Napp-Zinn K. (1985) *Arabidopsis thaliana*. In: Handbook of Flowering. Vol. I. Halevy H. A. (ed.) (Boca Raton, Florida: CRC Press) 492-503.
- Newman T., de Bruijn F. J., Green P., Keegstra K., Kende H., McIntosh L., Ohlrogge J., Raikel N., Somerville S., Thomashow M., Retzel E. and Somerville C. R. (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* **106** 1241-1255.
- Olszewski N. E., Martin F. B. and Ausubel F. M. (1988) Specialized binary vector for plant transformation: expression of the *Arabidopsis thaliana* AHAS gene in *Nicotiana tabacum*. *Nuc. Acids. Res.* **16** 10765-10782.
- Omichinski J. G., Clore G. M., Schaad O., Felsenfeld G., Trainor C., Appella E., Stahl S. J. and Gronenborn A. M. (1993) NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* **261** 438-446.
- Osada H., Grutz G., Axelson H., Forster A. and Rabbitts T. H. (1995) Association of erythroid transcription factors: complexes involving the LIM protein RBTN2 and the zinc-finger protein GATA1. *Proc. Natl. Acad. Sci. USA* **92** 9585-9589.
- Osborn T. C., Kole C., Parkin I. A. P., Sharpe A. G., Kuiper M., Lydiat D. J. and Trick M. (1997) Comparison of flowering time genes in *Brassica rapa*, *B. napus* and *Arabidopsis thaliana*. *Genetics* **146** 1123-1129.
- Park E.-C. and Horvitz R. (1986) Mutations with dominant effects on the behaviour and morphology of the nematode *Caenorhabditis elegans*. *Genetics* **113** 821-852.
- Parkin I. A. P., Sharpe A. G., Keith D. J. and Lydiat D. J. (1995) Identification of the A and C genomes of amphidiploid *Brassica napus* (oilseed rape). *Genome* **38** 1122-1131.
- Parks B. M. and Quail P. H. (1991) Phytochrome-deficient *hy-1* and *hy-2* long hypocotyl mutants of *Arabidopsis* are defective in phytochrome chromophore biosynthesis. *Plant Cell* **3** 1177-1186.
- Paz-Arez J., Ghosal D. and Saedler H. (1990) Molecular analysis of the C1-I allele from *Zea mays*: a dominant mutant of the regulatory C1 locus. *The EMBO J.* **9** 315-321.
- Pelham H. R. B. and Brown D. D. (1980) A specific transcription factor that can bind either the 5S RNA gene or 5S RNA. *Proc. Natl. Acad. Sci. USA* **77** 4170-4174.
- Peng J., Carol P., Richards D. E., King K. E., Cowling R. J., Murphy G. P. and Harberd N. P. (1997) The *Arabidopsis* *GAI* gene defines a signaling pathway that negatively regulates gibberellin responses. *Genes and Dev.* **11** 3194-3205.
- Peng J. and Harberd N. P. (1993) Derivative alleles of the *Arabidopsis* gibberellin-insensitive (*gai*) mutation confer a wild-type phenotype. *Plant Cell* **5** 351-360.
- Pepper A. E. and Chory J. (1997) Extragenic suppressors of the *Arabidopsis* *det1* mutant identify elements of flowering time and light response regulatory pathways. *Genetics* **145** 1125-1137.
- Pepper A. E., Delaney T., Washburn T., Poole D. and Chory J. (1994) *DET1*, a negative regulator of light-mediated development and gene expression in *Arabidopsis*, encodes a novel, nuclear-localized protein. *Cell* **78** 109-116.

- Ptashne M. (1988) How eukaryotic transcriptional activators work. *Nature* **335** 683-689.
- Putterill J. J., Ledger S. E., Lee K., Robson F., Murphy G. and Coupland G (1997) The flowering-time gene *CONSTANS* and homologue *CONSTANS-LIKE 1* (Accession No. Y10555 and Y10556) exist as a tandem repeat on chromosome 5 of *Arabidopsis* (PGR97-077). *Plant Physiol.* **114** 396.
- Putterill J. Robson F. Lee K. and Coupland G. (1993) Chromosome walking with YAC clones in *Arabidopsis*: isolation of 1700 kb of contiguous DNA on chromosome 5, including a 300 kb region containing the flowering-time gene *CO*. *Mol. Gen. Genet.* **239** 145-157.
- Putterill J., Robson F., Lee K., Simon R. and Coupland G. (1995) The *CONSTANS* gene of *Arabidopsis* promotes flowering and encodes a protein showing similarities to zinc finger transcription factors. *Cell* **80** 847-857.
- Quail P. H., Boylan M. T., Parks B. M., Short T. W., Xu Y. and Wagner D. (1995) Phytochromes: photosensory perception and signal transduction. *Science* **268** 675-680.
- Raikhel N. (1992) Nuclear targeting in plants. *Plant Physiol.* **100** 1627-1632.
- Raleigh E. and Wilson G. (1986) *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. *Proc. Natl. Acad. Sci. USA* **83** 9070-9074.
- Ramain P., Heitzler P., Haenlin M. and Simpson P. (1993) *pannier*, a negative regulator of *achete* and *scute* in *Drosophila*, encodes a zinc finger protein with homology to the vertebrate transcription factor GATA-1. *Development* **119** 1277-1291.
- Ray A., Lang D. J., Golden T. and Ray S. (1996) *SHORT INTEGUMENT (SINI)*, a gene required for ovule development in *Arabidopsis*, also controls flowering time. *Development* **122** 2631-2638.
- Rédei G. P. (1962) Supervital mutants of *Arabidopsis*. *Genetics* **47** 443-460.
- Rédei G. P. (1969) *Arabidopsis thaliana* (L.) Heynh. A review of the genetics and biology. *Biblio. Genet.* **21** 1-151.
- Rédei G. P. (1992) A heuristic glance at the past of *Arabidopsis* genetics. In: *Methods in Arabidopsis Research*. Koncz C., Chua N-H and Schell (eds.) World Scientific, Singapore, 1-15.
- Rédei G. P., Acedo G. and Gavazzi G. (1974) Flower differentiation in *Arabidopsis*. *Stadler Symp.* **6** 135-168.
- Rédei G. P. and Koncz C. (1992) Classical mutagenesis. In: *Methods in Arabidopsis Research*. Koncz C., Chua N-H and Schell (eds.) World Scientific, Singapore, 1-15.
- Reed J. W., Nagpal P., Poole D. S., Furuya M. and Chory J. (1993) Mutations in the gene for the red/far red light receptor phytochrome B alter cell elongation and physiological responses throughout *Arabidopsis* development. *Plant Cell* **5** 147-157.
- Reiter R.S., Williams J. G. K., Feldman K. A., Rafalski J. A., Tingey S. V. and Scolnik P. A. (1992) Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplified polymorphic DNAs. *Proc. Natl. Acad. Sci. USA* **89** 1477-1481.
- Rieping M., Fritz M., Prat S. and Gatz C. (1994) A dominant negative mutant of PG13 suppresses transcription from a cauliflower mosaic virus 35S truncated promoter in transgenic tobacco plants. *The Plant Cell* **6** 1087-1098.
- Robbins J., Dilworth S. M., Laskey R. A. and Dingwall C. (1991) Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64** 615-623.
- Robert L. S., Robson F., Sharpe A., Lydiate D. and Coupland G. (1998) Conserved structure and function of the *Arabidopsis* flowering time gene *CONSTANS* in *Brassica napus*. *Plant Mol. Biol.* In press.

- Rose A. B., Li J. and Last R. L. (1997) An allelic series of blue fluorescent *trp1* mutants of *Arabidopsis thaliana*. *Genetics* **145** 197-205.
- Rost B. (1996) PHD: predicting one dimensional protein structure by profile based neural networks. *Meth. in Enzym.* **266** 525-539.
- Ruiz-García L., Madueño F., Wilkinson M., Haughn G., Salinas J. and Martínez-Zapater J. M. (1997) Different roles of flowering-time genes in the activation of floral initiation genes in *Arabidopsis*. *Plant Cell* **9** 1921-1934.
- Rushton P. J., Macdonald H., Huttly A. K., Lazarus C. M. and Hooley R. (1995) Members of a new family of DNA-binding proteins bind to a conserved *cis*-element in the promoters of α -Amy2 genes. *Plant Mol. Biol.* **29** 691-702.
- Sadowski J., Gaubier P., Delseny M. and Quiros C. F. (1996) Genetic and physical mapping in *Brassica* diploid species of a gene cluster defined in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **251** 298-306.
- Sáez-Vásquez J., Raynal M., Meza-Basso L. and Delseny M. (1993) Two related, low-temperature-induced genes from *Brassica napus* are homologous to the human tumour *bbc1* (breast basic conserved) gene. *Plant Mol. Biol.* **23** 1211-1221.
- Saha S., Brickman J. M., Lehming N. and Ptashne M. (1993) New eukaryotic transcriptional repressors. *Nature* **363** 648-652.
- Saiki R. K., Gelfand D. H., Stoffel S., Scharf S. J., Higuchi R., Hom G. T., Mullis K. B. and Erlich H. A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239** 487-491.
- Sakamoto A., Minami M., Huh G. H. and Iwabuchi M. (1993) The putative zinc-finger protein WFZ1 interacts with a *cis*-acting element of wheat histone genes. *Eur. J. Biochem.* **217** 1049-1056.
- Sakai H., Medrano L. J. and Meyerowitz E. M. (1995) Role of *SUPERMAN* in maintaining *Arabidopsis* floral whorl boundaries. *Nature* **378** 199-203.
- Sánchez-García I. and Rabbitts T. H. (1994) The LIM domain: a new structural motif found in zinc-finger-like proteins. *TIG* **10** 315-320.
- Sanda S. L. and Amasino R. M. (1996a) Ecotype-specific gene expression of a flowering mutant phenotype in *Arabidopsis thaliana*. *Plant Physiol.* **111** 641-644.
- Sanda S. L. and Amasino R. M. (1996b) Interaction of *FLC* and late-flowering mutations in *Arabidopsis thaliana*. *Mol. Gen. Genet.* **251** 69-74.
- Sanger F., Nicklen S. and Coulson A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74** 5463-5467.
- Saurin A. J., Borden K. L. B., Boddy M. N. and Freemont P. S. (1996) Does this have a familiar RING? *TIBS* **21** 208-214.
- Schaffer R., Ramsey N., Samach A., Corden S., Putterill J., Carré I. A. and Coupland G. (1998) *LATE ELONGATED HYPOCOTYL*, an *Arabidopsis* gene encoding a MYB transcription factor, regulates circadian rhythmicity and photoperiodic responses. *Cell*. In press.
- Scheffler J. A., Sharpe A. G., Schmidt H., Sperling P., Parkin I. A. P., Luhrs W., Lydiat D. J. and Heinz E. (1997) Desaturase multigene families of *Brassica napus* arose through genome duplication. *Theor Appl. Genet.* **94** 583-591.
- Schindler U., Beckmann H. and Cashmore A. R. (1993) HAT3.1, a novel *Arabidopsis* homeodomain protein containing a conserved cysteine-rich region. *Plant J.* **4** 137-150.
- Schmeichel K. L. and Beckerle M. C. (1994) The LIM domain is a modular protein-binding interface. *Cell* **79** 211-219.
- Schmidt R., Love K., West J., Lenehan Z. and Dean C. (1997) Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. *Plant J.* **11** 563-572.

- Schmidt R., West J., Love K., Lenehan Z., Lister C., Thompson H., Bouchez D. and Dean C. (1995) Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science* **270** 480-483.
- Schmiedeskamp M. and Klevit R. E. (1994) Zinc finger diversity. *Curr. Op. Struc. Biol.* **4** 28-35.
- Schultz E. A. and Haughn G. W. (1991) *LEAFY*, a homeotic gene that regulates inflorescence development in *Arabidopsis*. *Plant Cell* **3** 771-781.
- Schultz E. A. and Haughn G. W. (1993) Genetic analysis of the floral initiation process (FLIP) in *Arabidopsis*. *Development* **119** 745-765.
- Schwechheimer C., Smith C. and Bevan M. W. (1998) The activities of acidic and glutamine-rich transcriptional activation domains in plant cells: design of modular transcription factors for high-level expression. *Plant Mol. Biol.* **36** 195-204.
- Shannon S. and Meeks-Wagner D. R. (1991) A mutation in the *Arabidopsis TFL1* gene affects inflorescence meristem development. *Plant Cell* **3** 877-892.
- Shannon S. and Meeks-Wagner D. R. (1993) Genetic interactions that regulate inflorescence development in *Arabidopsis*. *Plant Cell* **5** 639-655.
- Shirley B. W., Hanley S. and Goodman H. M. (1992) Effects of ionizing radiation on a plant genome: Analysis of two *Arabidopsis transparent testa* mutations. *Plant Cell* **4** 333-347.
- Simon M. C. (1995) Gotta have GATA. *Nature Genetics* **11** 9-11.
- Simon R. and Coupland G. (1996) *Arabidopsis* genes that regulate flowering time in response to day-length. *Sem. Cell. Dev. Biol.* **7** 419-425.
- Simon R., Igeño M. I. and Coupland G. (1996) Activation of floral meristem identity genes in *Arabidopsis*. *Nature* **384** 59-62.
- Smith H., Casal J. J. and Jackson G. M. (1990) Reflection signals and the perception by phytochrome of the proximity of neighbouring vegetation. *Plant Cell Environ.* **13** 73-78.
- Smith V. M., Lee P. P., Szychowski S. and Winoto A. (1995) GATA-3 dominant negative mutant. *J. Biol. Chem.* **270** 1515-1520.
- Snyder W. E. (1948) Mechanism of the photoperiodic response of *Plantago lanceolata*, a long day plant. *Am J. Bot.* **35** 520-525.
- Soppe W., Alonso-Blanco C., Koornneef M. and Peeters A. J. M. (1997) Cloning of *FWA* and characterisation of *EFS*, two genes involved in floral induction in *Arabidopsis*. *8th International Conference on Arabidopsis Research, Madison, Wisconsin*, abstr. 5-33.
- Southern E. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98** 503-517.
- Speulman E. and Salamini F. (1995) A barley cDNA clone with homology to the DNA-binding domain of the steroid hormone receptors. *Plant Sci.* **106** 91-98.
- Stafford H. A. (1990) *Flavonoid Metabolism*. Boca Raton, FL: CRC Press.
- Stiekema W. J., Heidekamp F., Dirkse W. G., van Beckum J., de Haan P., ten Bosch C. and Louwerse J. D. (1988) Molecular cloning and analysis of four potato tuber mRNAs. *Plant Mol. Biol.* **11** 255-269.
- Suárez-López P. and Coupland G. (1998) Plants see the blue light. *Science* **279** 1323-1324.
- Suge S. P. (1970) Changes in endogenous gibberellins in vernalized radish plants. *Plant Cell Physiol.* **11** 729-735.
- Sun T., Goodman H. M. and Ausubel F. M. (1992) Cloning the *Arabidopsis GAI* locus by genomic subtraction. *The Plant Cell* **4** 119-128.
- Sung Z. R., Belachew A., Shunong B. and Bertrand-Garcia R. (1992) *EMF*, an *Arabidopsis* gene required for vegetative shoot development. *Science* **258** 1645-1647.

- Tague B. W. and Goodman H. M. (1995) Characterization of a family of *Arabidopsis* zinc finger protein cDNAs. *Plant Mol. Biol.* **28** 267-279.
- Takasuji H., Mori M., Benfey P. N., Ren L. and Chua N.-H. (1994) Characterisation of a zinc finger DNA-binding protein expressed specifically in petunia petals and seedlings. *EMBO J.* **11** 241-249.
- Taylor W. R. (1986) The classification of amino acid conservation. *J. Theor. Biol.* **119** 205-218.
- Teakle G. R. and Gilmartin P. M. (1998a) Two forms of type IV zinc-finger motif and their kingdom specific distribution between the flora, fauna and fungi. *TIBS* **23** 100-102.
- Teakle G. R. and Gilmartin P. M. (1998b) Identification and characterization of a family of four GATA factor homologues from *Arabidopsis thaliana*. In press.
- Teutonico R. A. and Osborn T. C. (1994) Mapping RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage maps of *B. napus*, *B. oleracea*, and *Arabidopsis thaliana*. *Theor. Appl. Genet.* **89** 885-894.
- Thomas B. and Vince-Prue D. (eds.) (1997) Photoperiodic control of flower initiation. In: *Photoperiodism in Plants* (Academic Press) 1-256.
- Thurling N. and Depittayanan V. (1992) EMS induction of early flowering mutants in spring rape (*Brassica napus*). *Plant Breeding* **108** 177-184.
- Torii K. U., Mitsukawa N., Oosumi T., Matsuura Y., Yokoyama R., Whittier R. and Komeda Y. (1996) The *Arabidopsis* *ERECTA* gene encodes a putative receptor protein kinase with extracellular leucine-rich repeats. *Plant Cell* **8** 735-746.
- Trainor C. D., Evans T., Felsenfeld G. and Boguski M. S. (1990) Structure and evolution of a human erythroid transcription factor. *Nature* **343** 92-96.
- Tsang A. P., Visvader J. E., Turner C. A., Fujiwara Y., Yu C., Weiss M. J., Crossley M. and Orkin S. H. (1997) FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation. *Cell* **90** 109-119.
- Uberbacher E. C. and Mural R. J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88** 11261-11265.
- Unger E., Parsons R. L., Schmidt R. J., Bowen B. and Roth B. A. (1993) Dominant negative mutants of Opaque2 suppress transactivation of a 22-kD zein promoter by Opaque2 in maize endosperm cells. *The Plant Cell* **5** 831-841.
- Valvekens D., van Montague M. and van Lijsebettens M. (1988) *Agrobacterium tumefaciens*-mediated transformation of *Arabidopsis* root explants using kanamycin selection. *Proc. Natl. Acad. Sci. USA* **85** 5536-5540.
- Vitale N., Moss J. and Vaughan M. (1998) Molecular characterization of the GTPase-activating domain of ADP-ribosylation factor domain protein 1 (ARD1). *J. Biol. Chem.* **273** 2553-2560.
- Vizir I. Y., Anderson M. L., Wilson Z. A. and Mulligan B. J. (1994) Isolation of deficiencies in the *Arabidopsis* genome by γ -irradiation of pollen. *Genetics* **137** 1111-1119.
- Vollbrecht E., Veit B., Sinha N. and Hake S. (1991) The developmental gene *Knotted-1* is a member of a maize homeobox gene family. *Nature* **350** 241-243.
- von Arnim A. G. and Deng X.-W. (1993) RING-finger motif in *Arabidopsis thaliana* COP1 defines a new class of zinc-binding domain. *J. Biol. Chem.* **268** 19626-19631.
- Wagner D., Koloszarvari M. and Quail P. H. (1996) Two small spatially distinct regions of phytochrome B are required for efficient signaling rates. *The Plant Cell* **8** 859-871.

- Wahl G. M., Stem M. and Stark G. R. (1979) Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization using dextran sulphate. *Proc. Natl. Acad. Sci. USA* **76** 3683-3687.
- Wang Z.-Y., Green R., and Tobin E. (1997) The functions of the *CCA1* gene in the regulation of gene expression and plant development by light and the circadian clock. *8th International Conference on Arabidopsis Research, Madison, Wisconsin*, abstr 9-31.
- Wang Z.-Y., Kenigsbuch D., Sun L., Harel E., Ong M. S. and Tobin E. M. (1997) A Myb-related transcription factor is involved in the phytochrome regulation of an *Arabidopsis lhcb* gene. *Plant Cell* **9** 491-507.
- Ward E. R. and Jen G. C. (1990) Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.* **14** 561-568.
- Way J. C. and Chalfie M. (1988) *mec-3*, a homeobox-containing gene that specifies differentiation of the touch receptor neurons in *C. elegans*. *Cell* **54** 5-16.
- Weigel D., Alvarez J., Smyth D. R., Yanofsky M. F. and Meyerowitz E. M. (1992) *LEAFY* controls floral meristem identity in *Arabidopsis*. *Cell* **69**: 843-859.
- Weigel D. and Nilsson O. (1995) A developmental switch sufficient for flower initiation in diverse plants. *Nature* **377** 495-500.
- Weller J. L., Reid J. B., Taylor S. A. and Murfet I. C. (1997) The genetic control of flowering in pea. *TIPS* **2** 412-418.
- Wetmur J. G. and Davidson N. (1968) Kinetics of renaturation of DNA. *J. Mol. Biol.* **31** 349-370.
- Whitelam G. C. and Harberd N. P. (1994) Action and function of phytochrome family members revealed through the study of mutant and transgenic plants. *Plant Cell Environ.* **17** 615-625.
- Whitelam G. C., Johnson E., Peng J., Carol P., Anderson M. L., Cowl J. and Harberd N. (1993) Phytochrome A null mutants of *Arabidopsis* display a wild type phenotype in white light. *Plant Cell.* **5** 757-768.
- Wilkinson J. Q. and Crawford N. M. (1991) Identification of the *Arabidopsis CHL3* gene as the nitrate reductase structural gene *NIA2*. *The Plant Cell* **3** 461-471.
- Wilson R. N., Heckman J. W. and Somerville C. R. (1992) Gibberellin is required for flowering but not for senescence in *Arabidopsis thaliana* under short days. *Plant Physiol.* **100** 403-408.
- Winicov I. (1993) cDNA encoding putative zinc finger motifs from salt-tolerant alfalfa (*Medicago sativa* L.) cells. *Plant Physiol.* **102** 681-682.
- Yanagisawa S. (1995) A novel DNA-binding domain that may form a single zinc finger motif. *Nuc. Acids Res.* **23** 3403-3410.
- Yanagisawa S. (1996) A novel multigene family that the gene for a maize DNA-binding protein MNB1a belongs to: isolation of genomic clones from this family and some aspects of its molecular evolution. *Biochem. Mol. Biol. Int.* **38** 665-673.
- Yanagisawa S. (1997) Dof DNA-binding domains of plant transcription factors contribute to multiple protein-protein interactions. *Eur. J. Biochem.* **250** 403-410.
- Yang C., Chen L. and Sung Z. (1995) Genetic regulation of shoot development in *Arabidopsis*: role of the *EMF* genes. *Developmental Biology* **169** 421-435.
- Yang H.-Y. and Evans T. (1992) Distinct roles for the two cGATA-1 finger domains. *Mol. Cell. Biol.* **12** 4562-4570.
- Yanovsky M. J., Casal J. J. and Whitelam G. C. (1995) Phytochrome A, phytochrome B and HY4 are involved in hypocotyl growth responses to natural radiation in *Arabidopsis*: weak de-etiolation of the *phyA* mutant under dense canopies. *Plant Cell Environ.* **18** 788-794.

- Zachgo E. A., Wang M. L., Dewdney J., Bouchez D., Camilleri C., Belmonte S., Huang L., Dolan M. and Goodman H. M. (1996) A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6** 19-25.
- Zagotta M. T., Hicks K. A., Jacobs C. I., Young J. C., Hangharter R. P. and Meeks-Wagner D. R. (1996) The *Arabidopsis* *ELF3* gene regulates vegetative photomorphogenesis and the photoperiodic induction of flowering. *Plant J.* **10** 691-702.
- Zagotta M. T., Shannon S., Jacobs C. and Meeks-Wagner D. R. (1992) Early flowering mutants of *Arabidopsis thaliana*. *Aust. J. Plant Physiol.* **19** 411-418.
- Zambryski P., Depicker A., Kruger K. and Goodman H. (1982) Tumor induction by *Agrobacterium tumefaciens*: Analysis of the boundaries of T-DNA. *J. Mol. Appl. Genet.* **1** 361-370.
- Zhang B., Chen W., Foley R. C., Buttner M. and Singh K. (1995) Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequence. *Plant Cell* **7** 2241-2252.
- Zou J. and Taylor D. C. (1997) Cloning and molecular characterization of an *Arabidopsis thaliana* RING zinc finger gene expressed preferentially during seed development. *Gene* **196** 291-295.